HEART DISEASE PREDICTION USING DATA MINING

A PROJECT REPORT

Submitted by

V.KARTHIKA (810015104033)

M.SUHAILA (810015104090)

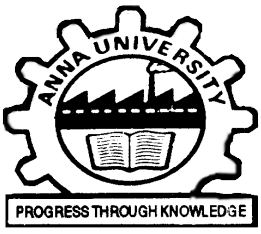
in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING



UNIVERSITY COLLEGE OF ENGINEERING (BIT CAMPUS)

TIRUCHIRAPALLI

ANNA UNIVERSITY CHENNAI-600 025

APRIL 2019

HEART DISEASE PREDICTION USING DATA MINING

A PROJECT REPORT

Submitted by

V.KARTHIKA (810015104033)

M.SUHAILA (810015104090)

Under the guidance of

Mrs.R.Sasikala

in partial fulfillment for the award of the degree

of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE AND ENGINEERING

UNIVERSITY COLLEGE OF ENGINEERING (BIT CAMPUS)

TIRUCHIRAPALLI

ANNA UNIVERSITY CHENNAI-600025

APRIL 2019

BONAFIDE CERTIFICATE

Certified that this project report"**HEART DISEASE PREDICTION USING DATA MINING"** is the bonafide work of **"KARTHIKA V (8100151014033)** and **SUHAILA M (810015104090)"** who carried out the project work under my supervision.

SIGNATURE

Mr.D.Vengatesan

HEAD OF THE DEPARTMENT

Assistant professor

Department of IT/CSE

Anna University(BIT CAMPUS)

Tiruchirapalli 620 024.

SIGNATURE

Mrs.R.Sasikala

SUPERVISOR

Department of CSE

University college of engineering

Anna University(BIT Campus)

Tiruchirapalli-620 024

Submitted for the ANNA UNIVERSITY Examination held on …….

INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

We hereby declare the work entitled "HEART DISEASE PREDICTION USING DATA MINING" is submitted in partial fulfillment of the requirement for the award of the degree in B.E., Computer Science and Engineering, University College of Engineering(BIT Campus), Tiruchirappalli, is a record of our own work carried out by us during the academic year 2018-2019 under the supervision and guidance of Mrs.R.Sasikala, Teaching Fellow, Department of Computer Science and Engineering, University College of Engineering(BIT Campus), Tiruchirappalli. The extent and source of information are derived from the existing literature and have been indicated through the dissertation at the appropriate places. The matter embodied in this work is original and has not been submitted for the award of any degree, either in this or any other University.

SIGNATURE OF THE CANDIDATES

V.KARTHIKA (810015104033)

M. SUHAILA (810015104090)

I certify that the declaration made above by the candidate is true.

SIGNATURE OF THE GUIDE

Mrs.R.SASIKALA

Teaching fellow,

Department of CSE,

University College of Engineering,

BIT Campus, Anna University,

Tiruchirappalli-620 024.

ACKNOWLEDGEMENT

I would like to convey my heartfelt thanks to our honorable Dean **Dr. T. SENTHILKUMAR,** Associate Professor for having provided me with all required facilities to complete my project without hurdles.

I would like to express my sincere thanks and deep sense of gratitude to guide **Mr. D. VENKATESAN,** Assistant Professor and Head, Department of Computer Science and Engineering, for his valuable guidance, suggestions and constant encouragement paved way for the successful completion of this project work.

I would like to thank my project guide **Mrs. R. SASIKALA,** Teaching Fellow, Department of Computer Science and Engineering, for his valuable guidance throughout the phase of the project. It is our responsibility to thank our project coordinator **Mr. C. SANKAR RAM,** Assistant Professor, Department of Computer science and Engineering for his constant inspiration that he has all through the project period.

I would like to thank **Mr. C. SURESH KUMAR,** Teaching Fellow, Department of Computer Science and Engineering, for his encouragement for this work.

I extend my thanks to all other teaching and non-teaching staffs for their encouragement and support.

I thank my beloved parents and friends, for their full support in my career development of this project.

ABSTRACT

Heart disease is a major life threatening disease that cause to death and it has a serious long term disability. There is wealth of data available within the health care system. However, there is lack of effective tools to discover hidden relationships and trends in data .Advanced data mining techniques can help remedial situations. This paper describes about a prototype using data mining techniques mainly Naïve Bayes and KNN. The physicians can also confirm their findings with the conformity of other physicians dealing with an identical case from all over the world. Medical diagnosis is considered as a significant yet intricate task that needs to be carried out precisely and efficiently. The automation of the same would be highly beneficial.

TABLE OF CONTENT

LIST OF ABBREVIATIONS

KNN K Nearest Neighbor

ASP Active Server Page

ADO ActiveX Data Object

ANN Artificial Neural Network

MSIL Microsoft Intermediate Language

MLPNN Multi Layer Perceptron Neural Network

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF THE PROJECT

Heart disease is the biggest cause of death nowadays. Blood pressure, cholesterol, pulse rate are the major reason for the heart disease. Some non-modifiable factors are also there. Such as smoking, drinking also reason for heart disease. The heart is an operating system of our human body. If the function of heart is not done properly means, it will affect other human body part also. Some risk factors of heart disease are Family history, High blood pressure, Cholesterol, Age, Poor diet, Smoking. When blood vessels are overstretched, the risk level of the blood vessels are increased. This leads to the blood pressure.

Blood pressure is typically measured in terms of systolic and diastolic. Systolic indicates the pressure in the arteries when the heart muscle contracts and diastolic indicates the pressure in the arteries when the heart muscle is in resting state. The level of lipids or fats increased in the blood are causes the heart disease. The lipids are in the arteries hence the arteries become narrow and blood flow is also become slow. Age is the non-modifiable risk factor which also a reason for heart disease. Smoking is the reason for 40% of the death of heart diseases. Because it limits the oxygen level in the blood then it damage and tighten the blood vessels. Various data mining techniques such as Naïve Bayes, KNN algorithm, Decision tree, Neural Network are used to predict the risk of heart disease . The KNN algorithm uses the K user defined value to find the values of the factors of heart disease. Decision tree algorithm is used to provide the classified report for the heart disease. The Naïve Bayes method is used to predict the heart disease through probability.

1.2 DATA MINING

Data Mining is about explaining the past and predicting the future by means of data analysis. Data mining is a multi-disciplinary field which combines statistics, machine learning, artificial intelligence and database technology. The value of data mining applications is often estimated to be very high. Many businesses have stored large amounts of data over years of operation, and data mining is able to extract very valuable knowledge from this data. The businesses are then able to leverage the extracted knowledge into more clients more sales, and greater profits. This is also true in the engineering and medical fields. Data mining predicts the future of modeling.

Predictive modelling is a process by which a model is created to predict outcome. If the outcome is categorical it is called categorical and if the outcome is numerical it is called regression. Descriptive modeling or clustering is assignment of observations into clusters so the observation of same cluster are beneficial.

Heart is vital part or an organ of the body. Life is subject to proficient working of heart. In the event that operation of heart is not proper, it will influence the other body parts of human, for example, mind, kidney, etc. Heart is simply a pump, which pumps the blood through the body. In the event that if blood in body is insufficient then many organs like cerebrum suffer and if heart quits working by, death happens inside minutes. Life is totally subject to effective working of the heart.The term Heart sickness alludes to illness of heart& vessel framework inside it.

There are number of elements which build the danger of Heart infection:

• family history of coronary illness

• smoking

• Poor eating methodology

• high pulse

• cholesterol

• high blood cholesterol

• obesity

• Physical inertia

1.3 Symptoms of a Heart Attack

Manifestations of a heart assault can include:

• Discomfort, weight, largeness, or agony in the midsection, arm, or beneath the breastbone.

• Discomfort emanating to the back, jaw, throat, or arm.

• Fullness, heartburn, or stifling feeling (may feel like indigestion).

• Sweating, queasiness, heaving, or unsteadiness.

• Extreme shortcoming, nervousness, or shortness of breath.

• Rapid or not regular heart beats

1.4 Types of heart Disease

A few sorts of Heart illnesses are

a). Coronary illness:

It otherwise called coronary supply route malady (CAD), it is the most wellknown kind of coronary illness over the world. It is condition in which plaque stores obstruct the coronary veins prompting a lessened supply of blood and oxygen to the heart.

b) Angina pectoris:

It is a therapeutic term for midsection torment that happens because of deficient supply of blood to the heart.

Otherwise called angina, it is a cautioning sign for heart assault. The midsection torment is at interims running for few seconds or minutes.

c). Congestive heart disappointment:

It is a condition where the heart can't pump enough blood to whatever is left of the body. It is generally known as heart disappointment.

d). Cardiomyopathy:

It is the debilitating of the heart muscle or a change in the structure of the muscle because of lacking heart pumping. A portion of the normal reasons for Cardiomyopathy are hypertension, liquor utilization, viral diseases,
and hereditary imperfections.

e). Innate coronary illness:

It alludes to the development of an irregular heart because of a deformity in the structure of the heart or its working. It is additionally a sort of innate ailment that kids are conceived with.

f). Arrhythmias:

It is connected with an issue in the musical development of the pulse. The pulse can be abating, quick, or unpredictable. These unusual heartbeats are brought about by a short out in the heart's electrical framework.

g). Myocarditis:

It is an aggravation of the heart muscle normally brought on by popular, parasitic, and bacterial contaminations influencing the heart.

It is an exceptional malady with few indications like joins agony, leg swelling or fever.

CHAPTER 2

LITERATURE SURVEY

Different types of studies have been done to focus on prediction of heart disease. various datamining techniques are used for diagnosis and achieved different accuracy level for different methods.

2.1 Web based health care detection using naïve bayes algorithm

Author: S.Indhumathi.etl

The preprocessed data has been considered as the training set. Two phase namely classification and prediction was discussed in that work. Preprocessing is done in the classification phase. The preprocessing includes cleaning of data, normalization and reduction of data, etc. In the prediction phase the disease types are classified and predicted, i.e. a training set is formed based on the disease type and the test set is formed based on the questions. The predicted results are sent to the doctor.

2.2 The prediction method for heart disease using Neural Network

Author: Chaitrali S.Dangare.etl

ANN, often just called a "neural network", is a mathematical model or computational model used for a biological purpose. In other words, it is an emulation of biological neural system. It has mainly three layers, i.e. the input layer, hidden layer and the output layer. The input is given to the input layer and the result is obtained in the output layer. Then the actual output and the expected output are compared. The back propagation has been applied to find the error and to adjust the weight between the output and the previous hidden layers. Once, the back propagation is completed, then the forward process is started and continued until the error is minimized.

2.3The risk level of heart disease prediction through hybrid algorithm Author: Shovon K. Pramanik. etl

Hybrid Algorithm is the combination of KNN algorithm and ID3. These algorithms are used for heart disease prediction. The KNN algorithm is used to preprocess the data; it is called as preprocessed algorithm. The preprocessed data are considered as training set and then the data has been classified into a tree structure. The ID3 algorithm is applied for the classifier to predict the heart disease. The incorrect values are classified through KNN algorithm.It is a non-parametric method which is used for classification and regression. Compared to other machine learning algorithm KNN is the simplest algorithm. This algorithm consist K-closet training examples in the feature space. In this algorithm K is a user defined constant. The test data are classified by assigning a constant value which is most chronic among the K-training samples nearest to the point.It shows the KNN has the strong consistency result.

2.4Association Rule for classification of Heart-attack patients

Author: N. Deepika et al

The extraction of significant patterns from the heart disease data warehouse was presented. The heart disease data warehouse contains the screening clinical data of heart patients. Initially, the data warehouse preprocessed to make the mining process more efficient. The first stage of Association Rule used preprocessing in order to handle missing values. Later applied equal interval binning with approximate values based on medical expert advice on Pima Indian heart attack data.

The significant items were calculated for all frequent patterns with the aid of the proposed approach. The frequent patterns with confidence greater than a predefined threshold were chosen and it was used in the design and development of the heart attack prediction system. The, Pima Indian Heart attack dataset used was obtained from the UCI machine learning repository. Characteristics of the patients like number of times of chest pain and age in years were recorded. The actions comprised in the preprocessing of a data set are the removal of duplicate records, normalizing the values used to represent information in the database, accounting for missing data points and removing unneeded data fields. Moreover it might be essential to combine the data so as to reduce the number of data sets besides minimizing the memory and processing resources required by the data mining.

2.5 K-means clustering with decision tree method to predict the heart disease

Author: Mai Shouman, et al.

They suggested several centroid selection methods for k-means clustering to increase efficiency. The 13 input attributes were collected from Cleveland Clinic Foundation Heart disease data set. The sensitivity, specificity, and accuracy are calculated with different initial centroids selection methods and different numbers of clusters. For the random attribute and random row methods, ten runs were executed and the average and best for each method were calculated. When comparing integrating k-means clustering and decision tree with traditional decision tree applied previously on the same data set, integrating k-means clustering with decision tree could enhance the accuracy of decision tree in diagnosing heart disease patients.

In Addition, integrating k-means clustering and decision tree could achieve higher accuracy than the paging algorithm in the diagnosis of heart disease patients. The accuracy achieved was 83.9% by the enabler method with two clusters.

2.6 Application of Data Mining Technique in Healthcare and Prediction of Heart Attacks

Author: K. Srinivas et al

The potential use of classification based data mining techniques such as Rule based, Decision tree, Naïve Bayes and Artificial Neural Network to the massive Volume of healthcare data. Tanagra data mining tool was used for exploratory data analysis, machine learning and statistical learning algorithms. The training data set consists of 3000 instances with 14 different attributes. The instances in the dataset are representing the results of different types of testing to predict the accuracy of heart disease. The performance of the classifiers is evaluated and their results are analyzed.

2.7Study of Heart Disease Prediction System using Data Mining Classification Techniques

Author: Dangare et al

Prediction System for heart disease used system contains huge amount of data, used to extract hidden information for making intelligent medical diagnosis. The main objective of this research was to build Intelligent Heart Disease Prediction System that gives diagnosis of heart disease using historical heart database. To develop the system, medical terms such as sex, blood pressure, and cholesterol like 13 input attributes are used. To get more appropriate results, two more attributes i.e. obesity and smoking, as attributes were considered as important attributes for heart disease.

A Multi-layer Perceptron Neural Networks (MLPNN) that maps a set of input data onto a set of appropriate. It consists of 3 layers input layer, hidden layer & output layer. There is connection between each layer & weights are assigned to each connection. The primary function of neurons of input layer is to divide input into neurons in hidden layer.

2.8 Association Rule mining based on the sequence number and clustering for heart attack prediction

Author:M A. Jabbar et al

The entire database is divided into partitions of equal size. The dataset with 14 attributes was used in work and also each cluster is considered one at a time for calculating frequent item. This approach reduces main memory requirement. To predict the heart attack in an efficient way the patterns are extracted from the database with significant weight calculation. The frequent patterns having a value greater than a predefined threshold were chosen for the valuable prediction of heart attack. Three mining goals were defined based on data exploration and all those models could answer complex queries in predicting heart attack.

2.9 Classify the Cardiovascular disease by using artificial neural network with back propagation error method

Author: Olatubosun Olabode et al

The Multi-layer perceptrons artificial neural networks with back-propagation error method were feed-forward nets with one or more layers of nodes between the input and output nodes. These additional layers contain hidden units or nodes that were not directly connected to both the input and output nodes. The neural network was trained using back propagation algorithm with sigmoid function on one hidden layer with the 16 input attributes.

Predictive models were used in variety of domains for the diagnosis. Dataset for this work were collected from federal medical fields. The input values obtained from the records of the forms the input variables in the input layer with 16 nodes. The neural network weights were initialized randomly. This work range of the weights was between [-0.5 and 0.5] and the learning rate was set between 0.1 and 0.9.

2.10 Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm

Author:M. Anbarasi et al

13 attributes involved in prediction of heart disease, proposed enhanced prediction of heart disease with feature subset selection using genetic algorithm using 10 attributes for predicting and data mining techniques after incorporating feature subset selection with high model construction time. Classification techniques are Naïve Bayes, Decision Tree and Classification by clustering.

The genetic search starts with zero attributes, and an initial population with randomly generated rules. Based on the idea of survival of the fittest, new population is constructed to comply with fittest rules in the current population, as well as offspring of these rules. Feature Extraction is the process of detecting and eliminating irrelevant, weakly relevant or redundant attributes or dimensions in a given data set.

CHAPTER 3

SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

- In existing system, the heart disease is predicted using naïve bayes algorithm.
- Naïve Bayes is the old method for classification and predictor selection.
- The Naïve Bayes model model is based on conditional independence model of each predictor given the target class.
- Heart disease is predicted using patients attribute using Naïve Bayes algorithm.

Drawbacks of the existing system

- There is a possibility for getting inaccurate results.
- User friendliness is very less.
- It consumes more time for processing the activities.

3.2 PROPOSED SYSTEM

- In Proposed system, we are using data mining techniques namely Naïve Bayes and KNN.
- We are using KNN as a new method for heart disease prediction and we are comparing it with other technique.
- By comparing it we can show that our proposed system is better
- than the other two techniques

ADVANTAGE:

- Better result accuracy.
- Reduced time complexity.
- The patient can be monitored from anywhere since it is a web application.

CHAPTER 4

SYSTEM SPECIFICATION

4.1 Hardware Requirements

Minimum Requirement

Monitors : 800x600 minimum resolutions at 256 colors minimum

Memory : Approximately 64 MB of on board memory

I/O : two or three button mouse and standard 101-key keyboard

MHZ : At least 166 MHZ processor

4.2 Software Requirements

Operating system -Windows XP Operating System

Front End -VB.NET

Back End -MS access

4.3 SOFTWARE DESCRIPTION

This chapter is about the software languages and the tools used in the development of the project.The software used here is Microsoft visual studio and the language used is .NET

. NET FRAMEWORK

The .NET Framework is an environment for building, deploying, and running Web Services and other applications. It is the infrastructure for the overall .NET Platform. The framework consists of three main parts: the Common Language Runtime, the class libraries, and ASP.NET.

The Common Language Runtime and class libraries, including Windows Forms, ADO.NET, and ASP.NET, combine together to provide services and solutions that can be easily integrated within and across a variety of systems. The .NET Framework provides a fully managed, protected, and feature-rich application execution environment, simplified development and deployment, and seamless integration with a wide variety of languages.

OVERVIEW OF .NET FRAMEWORK:

The .NET Framework is a managed, type-safe environment for application development and execution.

- It allocates memory for the storage of data and instructions
- Grants or denies the appropriate permissions to your application
- Initiates and manages application execution
- Manages the reallocation of memory for resources that are no longer needed

The .NET Framework consists of two main components

- The Common Language Runtime
- The .NET Framework class library.

COMMON LANGUAGE RUNTIME:

The common language runtime can be thought of as the environment that manages code execution. It provides core services, such as

- Code compilation Memory allocation
- Thread management, and
- Garbage collection
- .NET Framework class library
- The .NET Framework class library provides a collection of useful and reusable types that are designed to integrate with the common language runtime.
- Types are object-oriented and
- Fully extensible

Languages and the .NET Framework

- The .NET Framework is designed for cross-language compatibility.This means that .NET components can interact with each other no matter what language they were originally written in.
- This level of cross-language compatibility is possible because of the common language run time. When a .NET application is compiled, it is converted from the language it was written in (Visual Basic .NET, C#, or any other .NET compliant language) to *Microsoft Intermediate Language* (MSIL or IL). This is a low-level language designed to be read and understood by the common language run time. Because all .NET executables and DLLs exist as intermediate language, they can freely interoperate.

- The Common Language Specification defines the minimum standards that
- .NET language compilers must conform to, and thus ensures that any source code compiled by a .NET compiler can interoperate with the .NET Framework.
- The Common Type System ensures type compatibility between .NET components. Because .NET applications are converted to IL prior to deployment and execution, all primitive data types are represented as .NET types. Thus, a Visual Basic *Integer* and a C# *int* are both represented in IL code as a *System.Int32*. Because both languages use a common and inter conversable type system, it is possible to transfer data between components and avoid time-consuming conversions or hard-to-find errors.

The Structure of a .NET Application

The primary unit of a .NET application is the assembly. An assembly is a self-describing collection of code, resources, and metadata. The *assembly manifest* contains information about what is contained within the assembly. The assembly manifest provides

- Identity information, such as the name and version number of the assembly.
- A list of all types exposed by the assembly.
- A list of other assemblies required by the assembly.
- A list of code access security instructions for the assembly. This includes a list of permissions required by the assembly and permissions to be denied the assembly

Each assembly has one and only one assembly manifest, and it contains all the description information for the assembly.

The assembly manifest can be contained in its own separate file, or it can be contained within one of the assembly's modules.

Compilation and Execution of a .NET Application

When you compile a .NET application, it is not compiled to binary machine code; rather, it is converted to IL, which is a low-level set of instructions understood by the common language run time. This is the form that your deployed application takes—one or more assemblies consisting of executable files and DLL files in IL form. At least one of these assemblies will contain an executable file that has been designated as the entry point for the application.

When execution of your program begins,

1. The first assembly is loaded into memory.

At this point, the common language run time examines the assembly manifest and determines the requirements to run the program. It examines security permissions requested by the assembly and compares them to the system's security policy. If the system's security policy does not allow the requested permissions, the application will not be run.

2. If the application passes the system's security policy, the common language run time executes the code.

3. It creates a process for the application to run in and begins application execution.

4. When execution starts, the first bit of code that needs to be executed is loaded into memory and compiled into native binary code from IL by the common language run time's Just-In-Time (JIT) compiler.

- Once compiled, the code is executed and stored in memory as native code, so each portion of code is compiled only once during the execution of an application. Whenever program execution branches to code that has not yet been executed, the JIT compiler compiles it ahead of execution and stores it in memory as binary code.

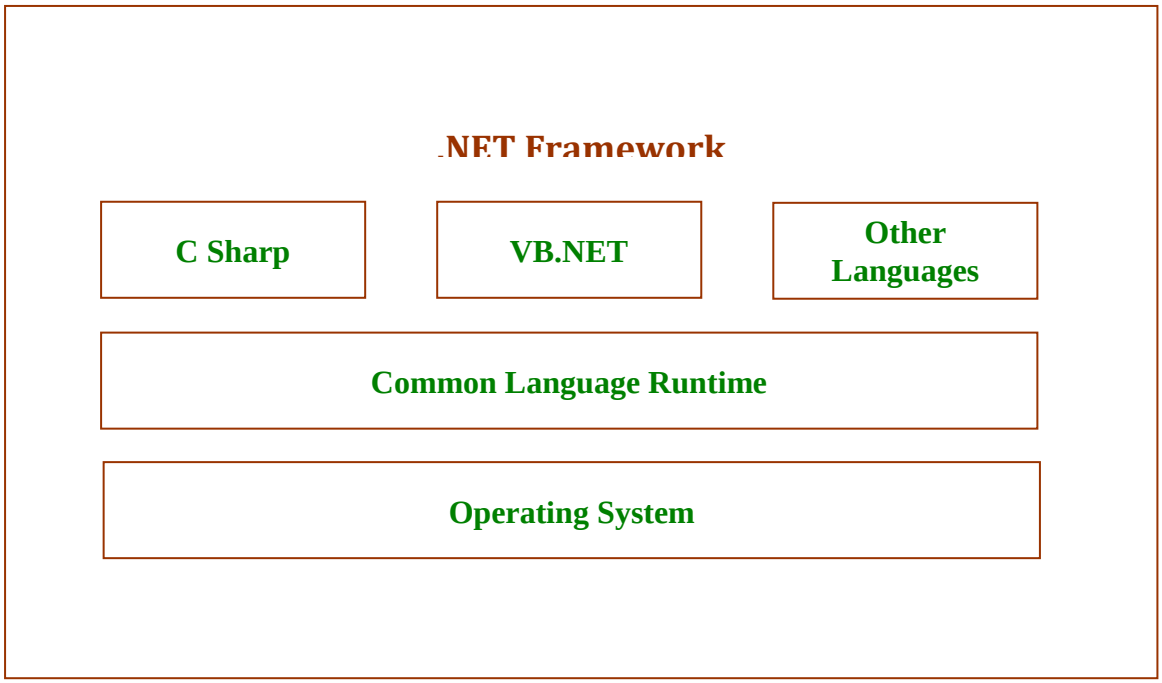


Fig 4.3 .NET Architecture

CHAPTER 5

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

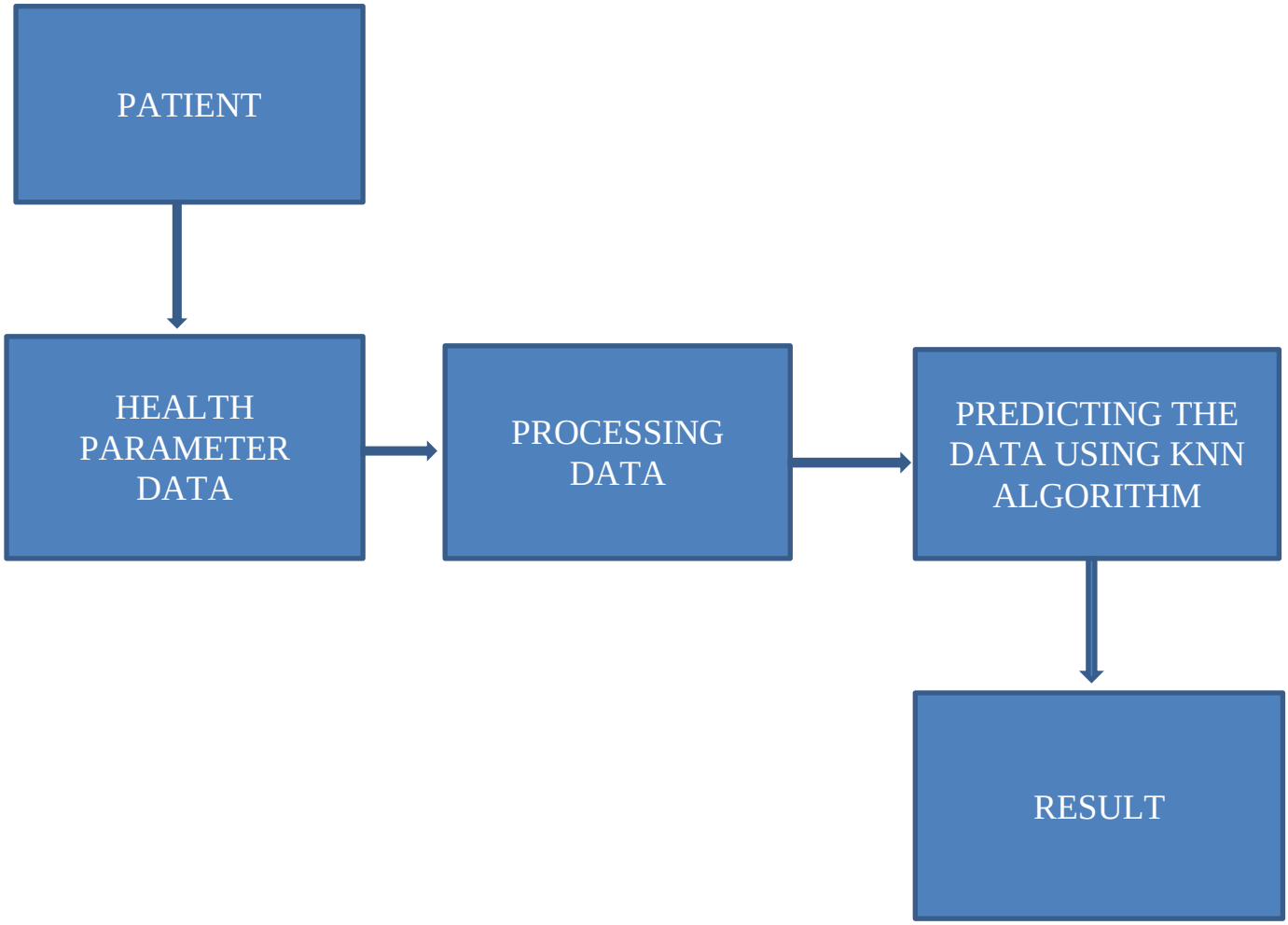


Fig 5.1 System architecture

5.2 MODULES:

- Registration and Login
- Patient Registration
- Patient details View
- Comparing KNN and Naïve Bayes algorithm

5.3 MODULES DESCRIPTION

5.3.1 Registration and Login

Admin can use this application by giving name and password. If name and password are existing the user is allowed to use this software. If not the system asks the admin level user to register first.

To become admin level user a person has to register with this system. User has to feed details such as name, email id, phone , address, password etc. These details are maintained in a database. While log in the name and password given by the user is compared with data in database and system decide whether user is allowed to tour the application

5.3.2 Patient Registration

This module maintains all the details such as patients name, phone no, email id and all heart related health parameters such as pulse, BP, frequency of heart, sugar levels etc., the phone numbers is used as key field using this field one can view patients details in future. These details are maintained in a database.

5.3.3 Patient details View

The details of a registered patient can be obtained using the phone number of the patient since the phone number is unique. By feeding just a phone number one can analyse the heart related decease if any .

The analyzing part analyzing the details using KNN and NAÏVE BAYES algorithms. The decease details are displayed on the screen .

Here in this part algorithms KNN and Naïve bayes are analysed separately**.**

5.3.4 Comparing KNN and Naïve Bayes algorithm

This part analyzing the details using KNN and NAÏVE bAYES algorithms. The result obtained by both the algorithm is compared .

The compared result always revealed that the KNN result is always superior, accurate and fast.

ALGORITHM

KNN

In recognition of patterns, the k-nearest neighbors algorithm (k-NN) is a method which is non-parametric and used for classification and regression. In both of the cases, the input consists of the k closest training examples in the feature space. The output of pattern recognition depends on whether k-NN is used for classification or regression:

• In **k-NN classification**, the output is a class membership.

An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors (k is a positive integer, typically small). If k = 1, then the object is simply assigned to the class of that single nearest neighbor.

• In **k-NN regression**, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

- K-NN is a type of instance-based learning, or lazy learning, where the function is only approximated locally and all computation is deferred until classification.
- *K*-NN algorithm among the simplest of all machine learning algorithms.
- Both for classification and regression, it can be useful to assign weight to the contributions of the neighbors, so that the nearer neighbors contribute more to the average than the more distant ones. For example, a common weighting scheme consists in giving each neighbor a weight of 1/*d*, where *d* is the distance to the neighbor.
- The neighbors are taken from a set of objects for which the class (for *k*-NN classification) or the object property value (for *k*-NN regression) is known.

- This can be thought of as the training set for the algorithm, though no explicit training step is required. A shortcoming of the *k*-NN algorithm is that it is sensitive to the local structure of the data. The algorithm is not to be confused with *k*-means, another popular machine learning technique.

The Naive Thomas Bayes rule relies on Bayesian theorem as given by on top of equation Steps in rule area unit as follows:

1. Every information sample is diagrammatical by associate n dimensional feature vector, X = (x1, x2.. xn), portraying n measurements created on the sample from n attributes, severally A1, A2, An.

2. Suppose that there are a unit m categories, C1, C2Cm. Given associate unknown information sample, X (i.e., having no category label), the category .

The category can predict that X belongs to the class having the very best posterior chance, conditioned if and solely if: P(Ci/X)>P(Cj/X) for all 1< = j< = m and j!= I therefore we tend to maximize P(Ci|X). The category Ci that P(Ci|X) is maximized is called the most posteriori hypothesis by Thomas Bayes theorem.

3. As P(X) is constant for all categories, solely P(X|Ci)P(Ci) want be maximized. If the category previous chances don't look like to be proverbial, then it's ordinarily considered that the categories area unit equally possible, i.e. P(C1) = P(C2) = ..= P(Cm),and we would thus maximize P(X|Ci). Otherwise, we tend to maximize P(X|Ci)P(Ci). Note that the category previous chances could also be calculable

by P(Ci) = si/s ,where Si is that the range of coaching samples of sophistication Ci, and s is that the total range of coaching samples, On X. That is, the naive chance assigns associate unknown sample X to the category Ci.

5.4 FLOW CHART

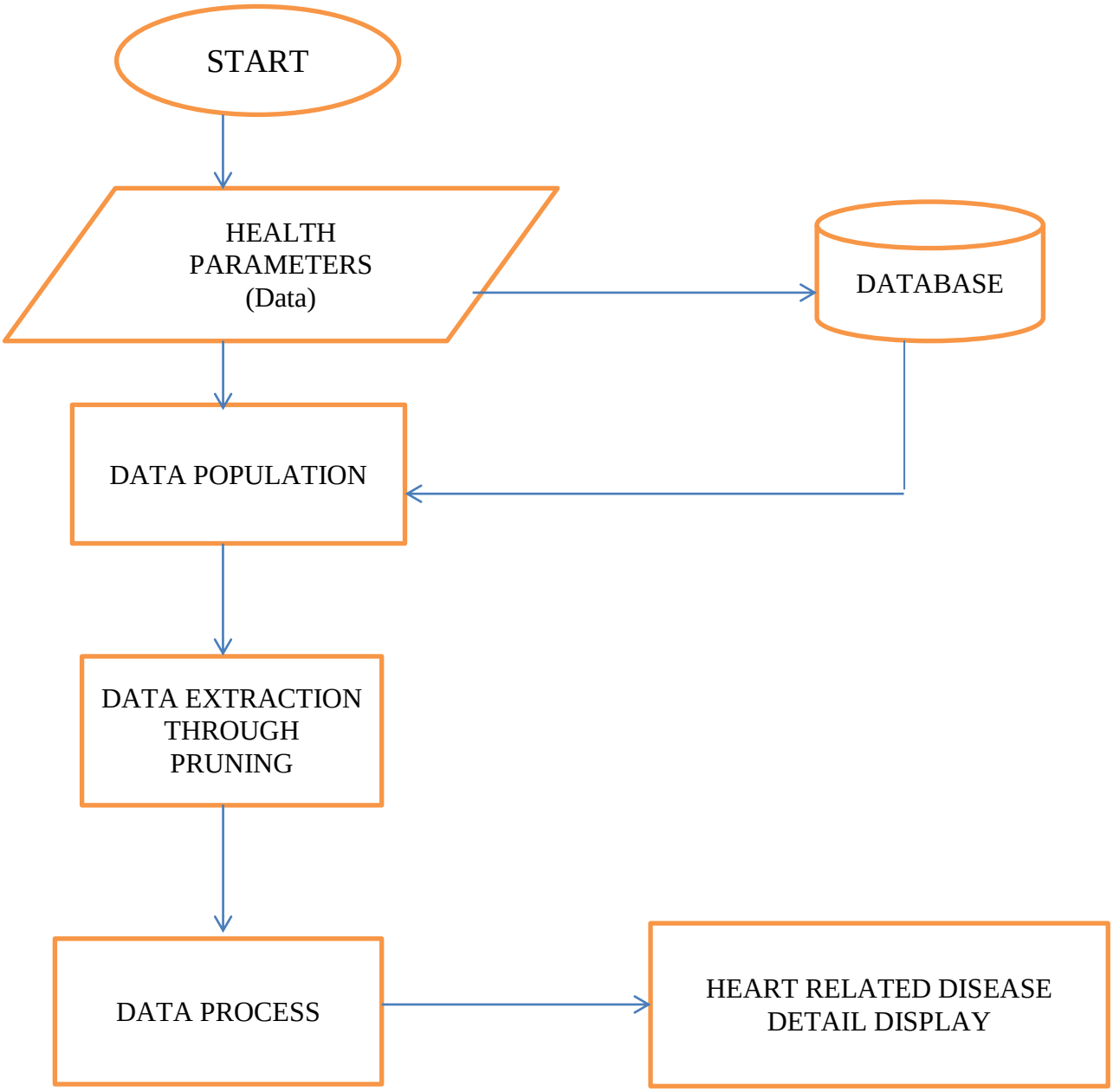


Fig.5.4. Flow chart of risk level prediction.

5.5 UML DIAGRAM

5.5.1 USE CASE DIAGRAM

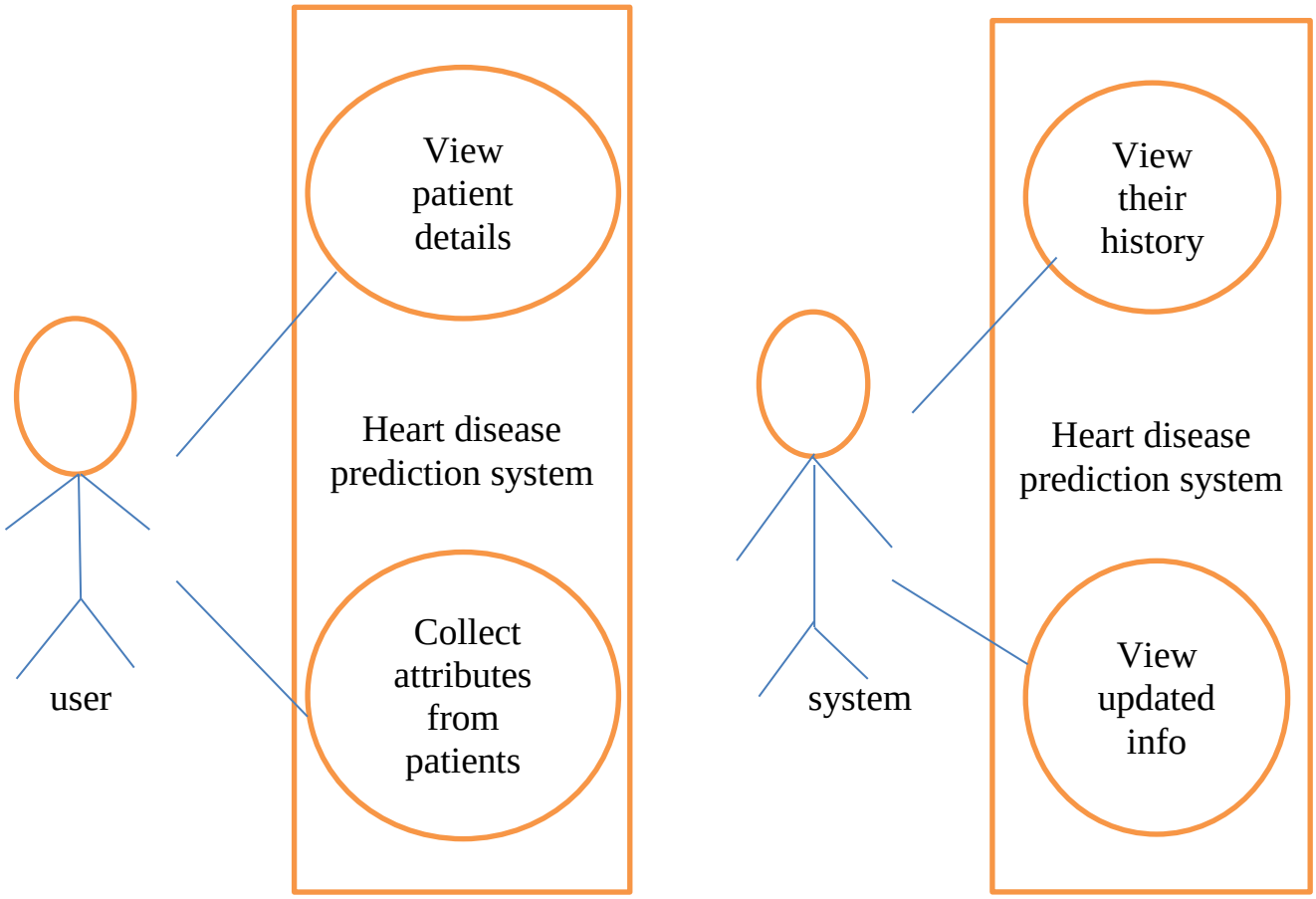


Fig 5.5.1. Use case diagram

A use case diagram at its **s**implest is a representation of a user's interaction with the system that shows the relationship between the user and the different use casesin which the user is involed.It can identify the different types of users of a system and the different use cases and will often be accompanied by other types of diagrams as well.

5.5.2 ACTIVITY DIAGRAM

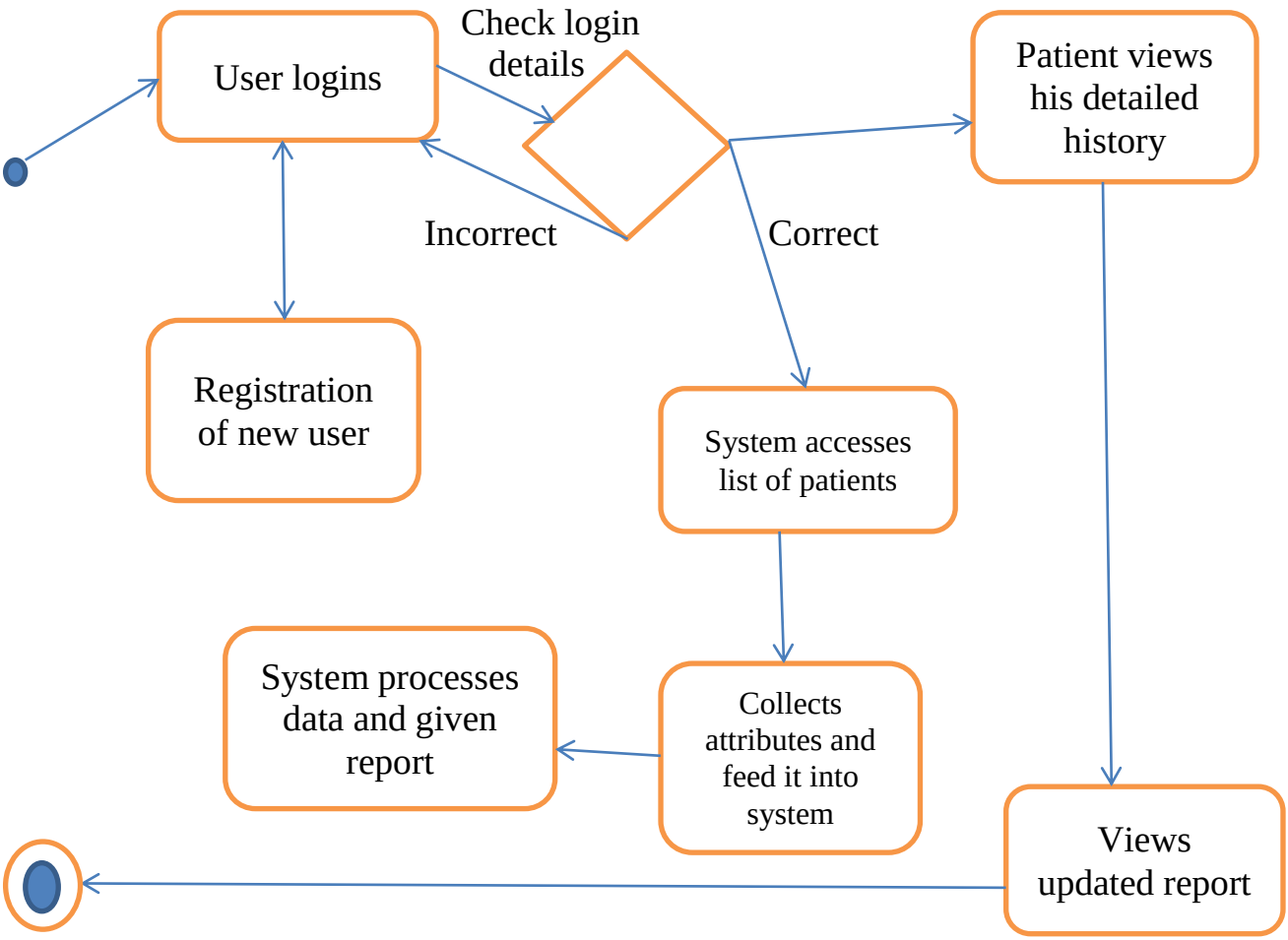


Fig 5.5.2. Activity Diagram

Activity diagram is another important diagram in UML to describe the dynamic aspects if the system .It is basically a flowchart to represent the flow from one activity to another activity .The activity can be described as an operation of the system .The control flow from is drawn from one operation to another .This flow can be sequential or concurrent. Activity diagrams deal with all types of flow control by using different elements such as fork, join …etc

5.5.3 FUNCTIONAL BLOCK DIAGRAM

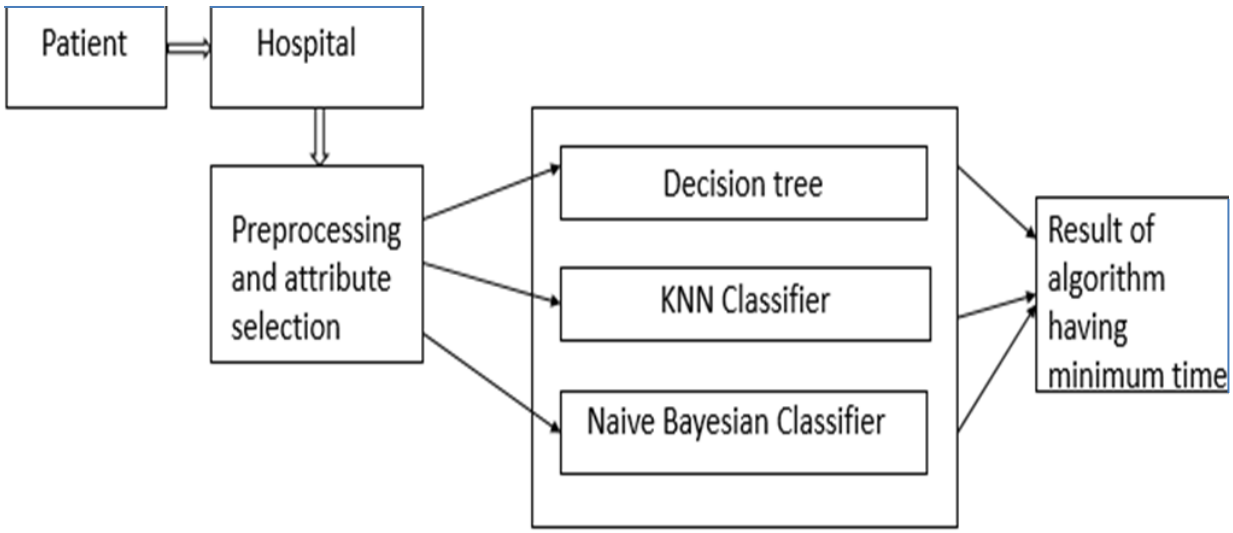


Fig.5.5.3. Functional Block Diagram

A Functional block diagram in systems engineering and software engineering is a block diagram.It describes the function and inter relationships of a system.

The functional block diagram can picture:

- The relationship between the functions,and
- The functional sequences and paths for matter and or signals

CHAPTER 6

TESTING

6.1 SYSTEM TESTING

Testing is the major control measure used during software development. Its basic function is to detect errors in the software. During requirement analysis and design, the output is a document that is usually textual and no executable.

After the coding phase, computer programs are available that can be executed for testing purpose. This implies that testing not only has to uncover errors introduced during previous phase. Thus the goal of testing is to uncover the requirements, design and coding errors in the programs.

System testing is an important stage in the system development life cycle. Even though it is an expensive process, it is essential for thorough checking and fault finding for different sets of conditions.

The common view of testing held by users is performed to prove that the software is free from bugs. There are many ways of testing the reliability, completeness, robustness and maintainability of a software package.

6.2 FUNCTIONAL TESTING

This testing phase includes checking the functionality of the system we have developed. This may include checking the overall functionality and to check the each module one by one.

The basis for deciding the test cases in this is the requirement or the specification of the program or the module. For the entire system the test cases are designed from the requirement specification document for the system.

Test cases are given for testing against requirements of the unit being tested. The unit modifies a test databases for the integrity of the databases for the integrity of the databases after the operation. Test cases for data flow coverage.

Test cases based on experience such as testing for boundary conditions minimum, maximum and off by one boundary.

1) Equivalence Class Partitioning –

Here we divide the domain of all inputs into set of equivalence classes so that any test in the equivalence class succeeds then every test in that class will succeed.

2) Cause Effect Graphing –

This technique starts with identifying causes and effects of the system under testing. A cause is a distinct input condition. Each condition forms a node in the cause effect graph.

3) Boundary Value Analysis –

It is observed that programs work correctly for a set of values in equivalence class fail on some specific values. These values often lie on boundary of equivalence class.

6.3 STRUCTURAL TESTING

This testing is concerned with the implementation of the program. In this we have to exercise the different programming structures and data structures used in the program. We have tested that whether this project can be implemented or not.

Test Case Generation and Tool Support

- To generate test cases tools are not that easily available and due to nature of the program a fully automated tool for selecting the test cases to satisfy the criterion is generally not possible hence tools can, at best, aid the tester.
- One method for generating the test cases to randomly select test cases until the desired criterion is satisfied. This can result in lot of redundant test cases, as many test cases will exercise the same paths. The usual procedure in testing is to create test data for initial test and to use live data for later testing.
- Test data cover all cases and should be exhaustive. Artificial data are created only for testing. It is generated to test all the combinations of formats and values. Moreover, newly designed system can be tested initially by giving the test data rather than live data since live data is voluminous and not available always.
- For unit testing, structural testing based on branch coverage criteria will be used.

6.4 LEVELS OF TESTING

6.4.1 Unit Testing

There are several levels of testing. These several levels attempt to detect different types of faults. The first level of testing is called **Unit Testing**. In this different modules are tested against the specification produced during the design of the modules.

Unit testing is essentially for verification of the code produced during the coding phase and hence the goal is to test the internal logic of the modules. It is typically done by the programmer of the module.

A module is considered for integration and used by others after it has been unit tested satisfactorily.

6.4.2 INTEGRATION TESTING

The Next level of testing is **Integration Testing**. In this many unit tested modules are combined into subsystems which are then tested. The goal here is to see if modules can be integrated properly.

The emphasis is on testing the interface between modules.Strategies for integrating software component into a functioning product include the

1. Bottom up Integration
2. Top Down Integration
3. Sandwich Integration.

The next levels are System Testing , It has an important role in software development.There are 2 following methods ,

- volume Testing
- Recovery Testing

6.4.3 RECOVERY TESTING

Recovery testing is a forced system failure is induced to test the back up recovery procedure .

There are 3 types of data namely

- Sample Data
- Live Data

- Trial Data

Acceptance Testing is sometimes performed with the realistic data of the client to demonstrate that software is working satisfactorily.

These levels of testing are performed when system is being built from components that have been coded.

After completing the project we have tested with details which are collected from many patients. The data records contains phone email id, and health parameters such as heart Beat, pressure, dizziness, O2 level, sugar etc.,The produced result by software is compared with manually calculated values. The result by the computer is more accurate and fast.

The links between pages also tested and all links or worked properly starting and ending point of the project are at login page. So this package can handled by the semiskilled employees.

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 CONCLUSION

A conclusion is often created that KNN is best among all the classification techniques after we name prediction or classification of a nonlinear knowledge. KNN uses nearest neighbor to find the optimal solution. Previous version of KNN are limited for the small size of datasets But the nearest neighbor is very handy and useful for all kind of datasets even for large size datasets. The solution found using this technique is very accurate. It is also faster the other two techniques we have mentioned. It has strong consistency results.

7.2 FUTURE WORK

In future the application can be modified as embedded application by embedding the electronic components with this system to monitor the patients through online.Since it is an web application, one can view the result from anywhere in the world.

APPENDIX 1

SAMPLE SOURCE CODE

```
Imports Microsoft.VisualBasic
Public Class knnalgarithm
Dim heart As Integer
Dim bp As Integer
Dim cholestral As Integer
Dim dia As Integer
Dim freq As Integer
Dim flutter As Integer
Dim bloodsup As Integer
Dim oxi As Integer
Dim wt_primary As Integer
Dim cor As Integer
Dim ill As Integer
Dim bloodalco As Integer
Dim beatshort As Integer
Dim lessfre As Integer
Dim swell As Integer
Dim caugh As Integer
Dim pressure As Integer
Dim dizz As Integer
Dim gastro As Integer
Dim virus As Integer
```

```
Dim pain As Integer
Dim sweat As Integer
Function pridiction_coronary(ByVal pwt_primary As Integer, ByVal pbloodsup
As Integer, ByVal poxi As Integer) As Integer
oxi = poxi
bloodsup = pbloodsup
cor = 0
ill = 0
If ((pwt_primary >= 100) And (oxi > 0) And (bloodsup > 0)) Then
ill = 1
End If
Return pwt_primary
End Function
Function anginopectories(ByVal pbeatshort As Integer, ByVal pbloodsup As
Integer, ByVal poxi As Integer) As Integer
beatshort = pbeatshort
oxi = poxi
bloodsup = pbloodsup
If (wt_primary >= 100) And (oxi > 0) And (bloodsup > 0) And (beatshort > 0)
Then
ill = 6
End If
If (wt_primary < 100) And (oxi > 0) And (bloodsup > 0) And (beatshort > 0)
Then
ill = 5
End If
Return ill
 End Function
```

```
Function congestivedisappointment(ByVal pbloodsup As Integer, ByVal palchohol As Integer) As Integer
bloodsup = pbloodsup
bloodalco = palchohol
If (bloodsup > 0) And (bloodalco > 0) Then
ill = 7
End If
Return ill
End Function
Function cardiomyopathy(ByVal pvirus As Integer, ByVal pdizz As Integer, ByVal plessfre As Integer, ByVal pswell As Integer, ByVal pcaugh As Integer, ByVal ppressure As Integer, ByVal dizzy As Integer, ByVal pgas As Integer) As Integer
gastro = pgas
lessfre = plessfre
swell = pswell
caugh = pcaugh
pressure = ppressure
dizz = pdizz
virus = pvirus
Dim wt_car As Integer
wt_car = gastro + lessfre + pcaugh + pressure + pdizz + virus
If ((lessfre > 0) And swell > 0 And caugh > 0 And pressure > 0 And dizzy > 0 And gastro >= 0) Then
ill = 8
End If
```

```
If ((virus) And swell > 0 And caugh > 0 And pressure > 0 And dizzy > 0 And
gastro >= 0) Then
ill = 9
End If
If (wt_car > 200) Then
ill = 10
End If
Return ill
End Function
Function arrhythmia(ByVal pfluttering As Integer, ByRef pheartbeat As
Integer, ByRef chestpain As Integer, ByRef pshortness As Integer, ByRef pdizz
As Integer, ByRef psweat As Integer, ByRef pfreq As Integer) As Integer
flutter = pfluttering
heart = pheartbeat
pain = chestpain
beatshort = pshortness
dizz = pdizz
sweat = psweat
freq = pfreq
Dim tot As Integer
'Dim flutterwt As Integer
tot = flutter + heart + pain + beatshort + sweat + freq
If (tot >= 250) Then
ill = 14
End If
If (flutter > 0 And heart > 0 And pain > 0 And beatshort > 0 And dizz > 0 And
sweat > 0 And freq > 0) Then
ill = 11
```

```
End If
' If (tot >= 100 And (flutter > 0 Or heart > 0 And pain > 0 And beatshort > 0 And dizz > 0 And sweat > 0 And freq > 0)) Then
'ill = 12
'End If
'If (flutter > 0 And heart > 0 Or pain > 0 Or beatshort > 0 Or dizz > 0 Or sweat > 0 Or freq > 0) Then
'ill = 13
'End If
Return ill
End Function
Function diseasetobedisplayed(ByVal primary As Integer, ByVal cor As Integer, ByVal agni As Integer, ByVal cong As Integer, ByVal coria As Integer, ByVal arri As Integer) As String
Dim ret As String
ret = " Report"
Dim arriflag As Integer
arriflag = 0
If (arri = 11) Then
ret = ret + "   Severe arrithmia is detected "
arriflag = arriflag + 1
End If
If (arri = 12) Then
ret = ret + "   Moderate arrithmia is detected "
arriflag = arriflag + 1
End If
If (arri = 13) Then
ret = ret + "  less Severe arrithmia is detected "
```

```
arriflag = arriflag + 1
End If
If (arri = 14) Then
ret = ret + "   Mild arrithmia is detected "
arriflag = arriflag + 1
End If
Dim coriaflag As Integer
coriaflag = 0
If (coria = 8) Then
ret = ret + "  Severe  Coriomyopathy detected "
coriaflag = coriaflag + 1
End If
If (coria = 9) Then
ret = ret + "  Severe Coriomyopathy detected "
coriaflag = coriaflag + 1
End If
If (coria = 10) Then
ret = ret + "   Moderate Cariomyopathy  detected "
coriaflag = coriaflag + 1
End If
Dim agniflag As Integer
agniflag = 0
If (agni = 6) Then
ret = ret + "  Severe Angino Pectories illness detected "
agniflag = agniflag + 1
End If
If (agni = 5) Then
ret = ret + "  Moderate Angino Pectories illness detected "
```

```
agniflag = agniflag + 1
End If
If (agni = 4) Then
ret = ret + "   more than moderate Angino Pectories illness detected "
agniflag = agniflag + 1
End If
Dim congflag As Integer
congflag = 0
If (cong = 7) Then
ret = ret + "  Congestive Heart Disappointment detected "
congflag = congflag + 1
End If
If (primary >= 100) Then
ret = ret + "  Mild Coronary illness detected "
congflag = congflag + 1
End If
Dim corflag As Integer
corflag = 0
If (cor = 1) Then
ret = ret + "  Severe Coronary illness detected "
corflag = corflag + 1
End If
If (cor = 2) Then
ret = ret + "  Moderate Coronary illness detected "
corflag = corflag + 1
End If
Dim totflag As Integer
totflag = arriflag + coriaflag + agniflag + congflag + corflag
```

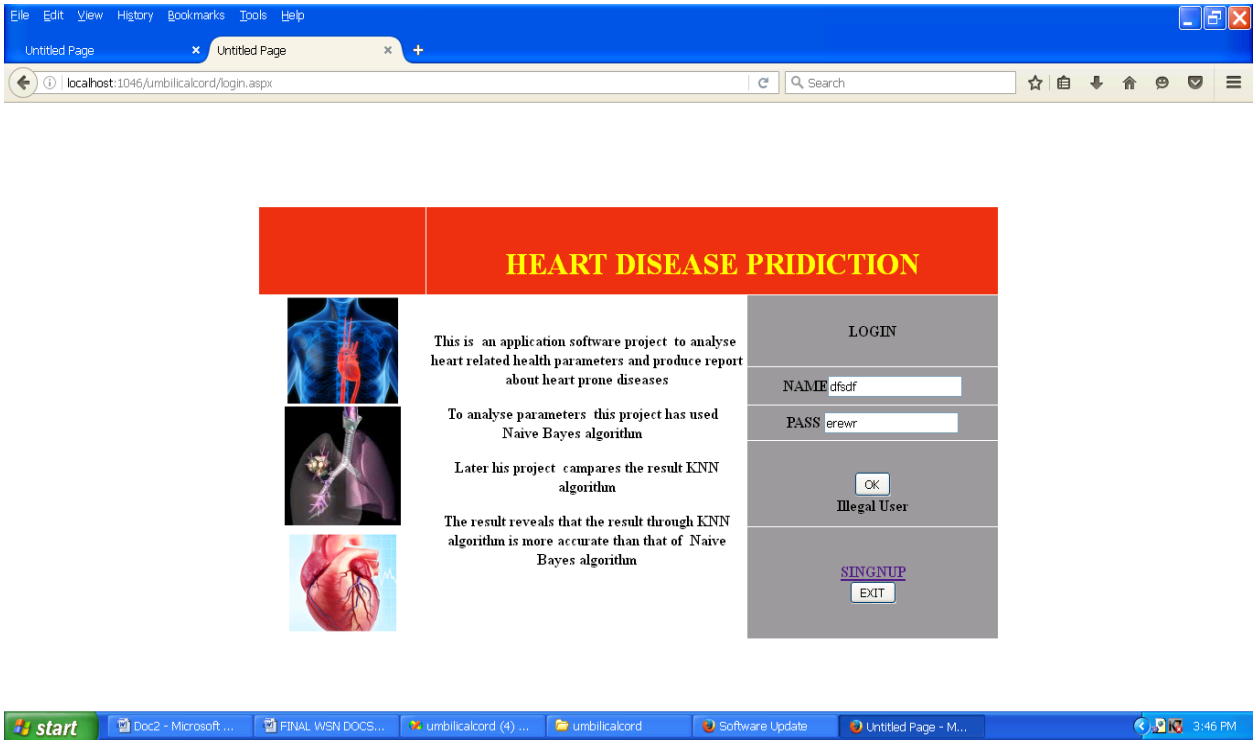
```
If (totflag > 1) Then
ret = ret + "multile Heart prone deseases attacked"
End If
Return ret
End Function
Function primary_diagoizedreport(ByVal a As Integer) As Integer
Return a
End Function
Function weight_entropy_primary_investigation(ByVal heart As Integer, ByVal
bp As Integer, ByVal cholestral As Integer) As Integer

wt_primary = heart + bp + cholestral
wt_primary = primary_diagoizedreport(wt_primary)
Return wt_primary
End Function
' MsgBox("Time consuming" + CType(counter, String),
MsgBoxStyle.Critical, "Time Taken")
'tim_taken_new = speed2.new_processing_speed
'Return tim_taken_new
End Function
```

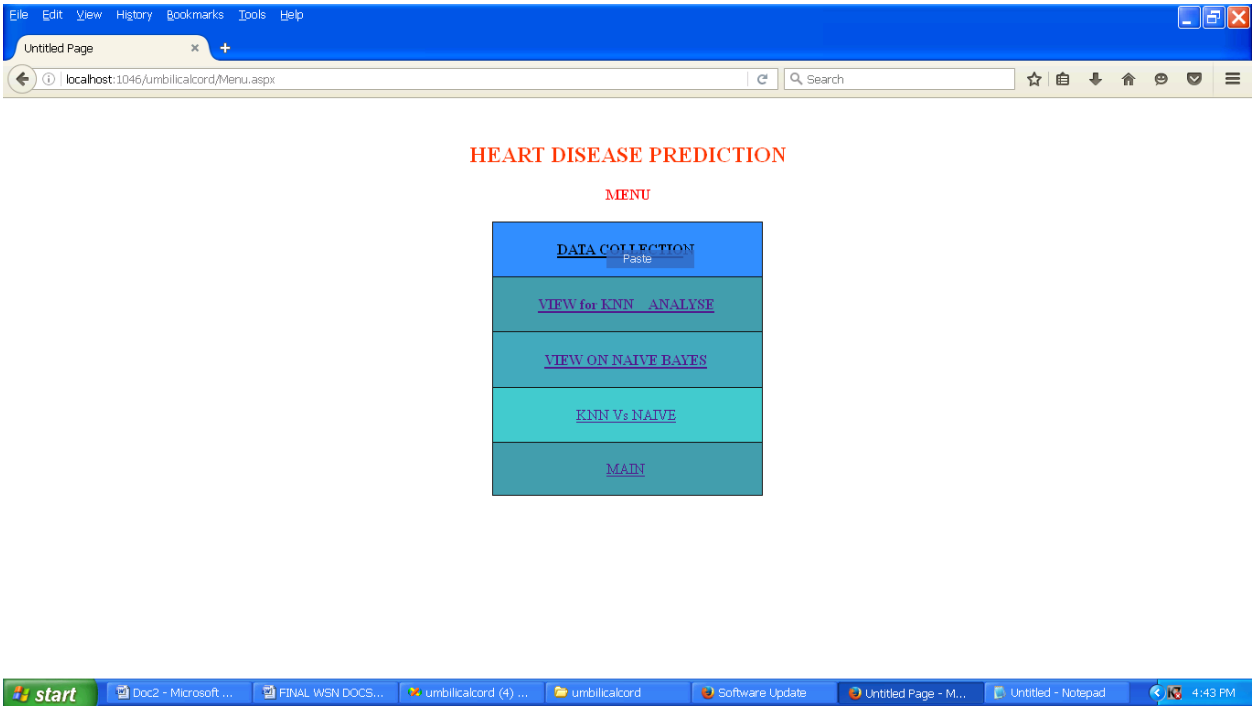
APPENDIX 2

SCREENSHOTS

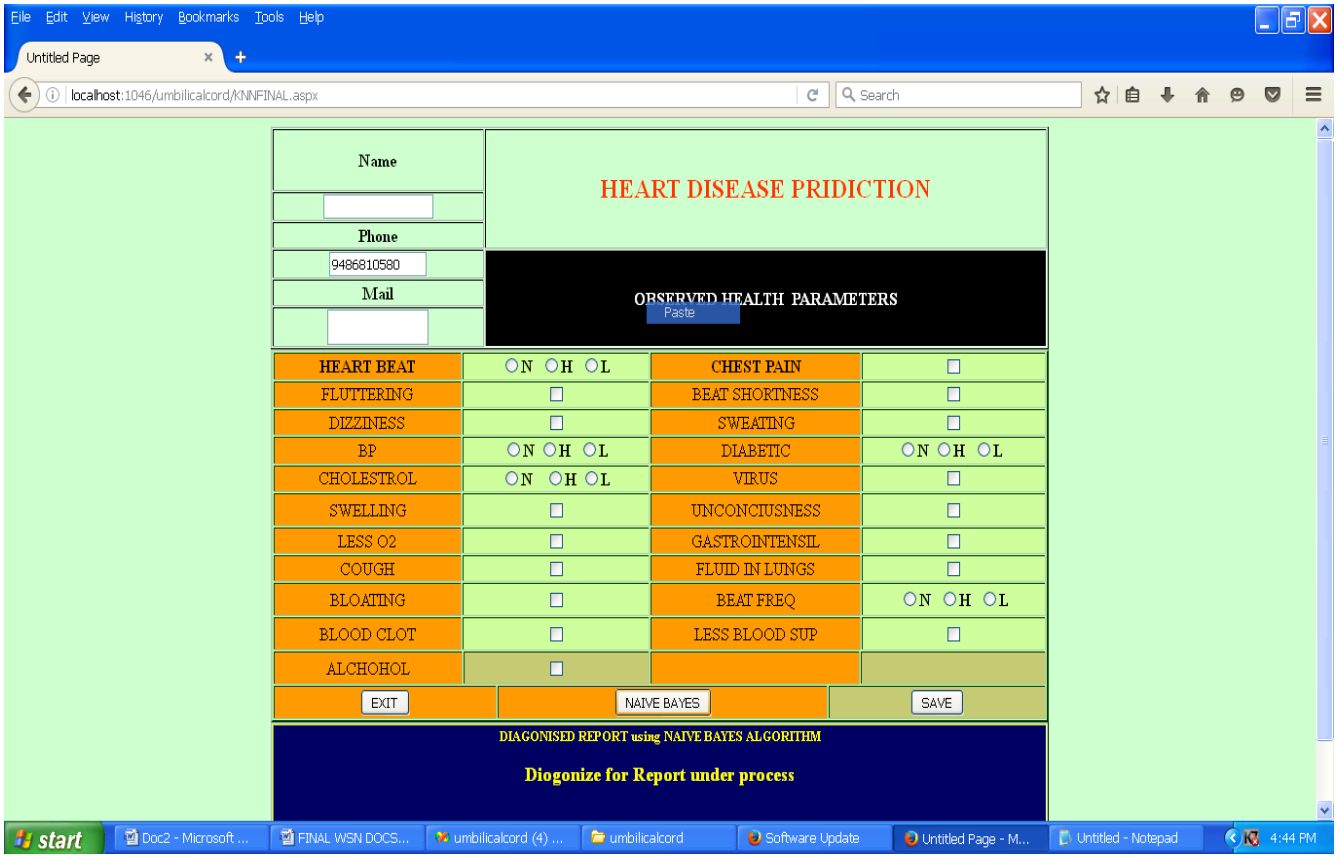
SCREENSHOT 1: LOGIN PAGE



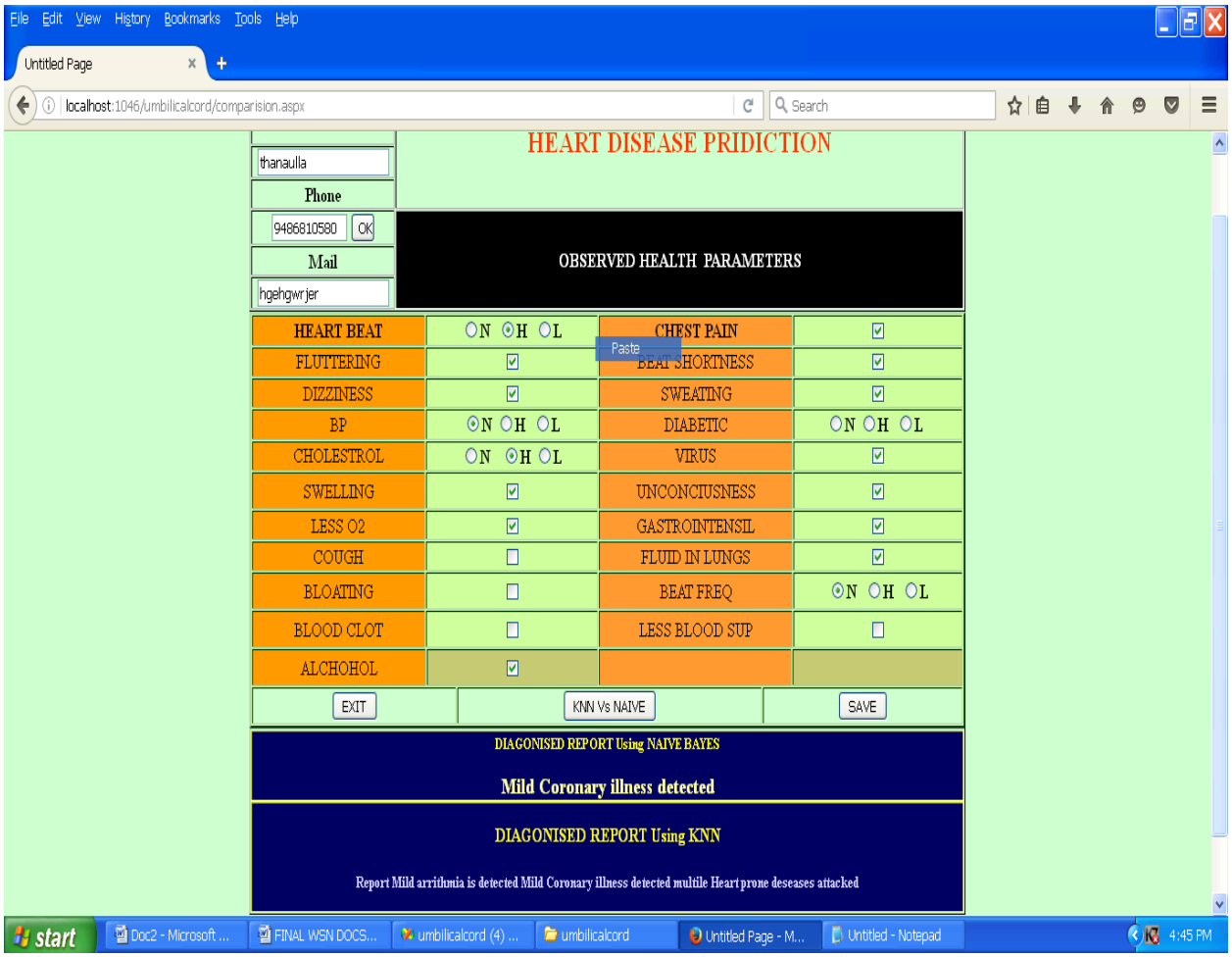
SCREENSHOT 2: Menu page



SCREEN SHOT 3: Prediction using naïve bayes



SCREEN SHOT 4: Prediction using KNN



REFERENCES

[1] S.Indhumathi, Mr.G.Vijaybaskar, "Web based health care detection using naive Bayes algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET**)**, Volume 4 Issue 9, pp.3532-36, September 2015.

[2]Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A Data mining approach for prediction of heart disease using neural network's", International Journal of Computer Engineering & Technology(IJCET)**)**, Volume 3, Issue 3, October – December (2012), pp. 30-40.

[3] Shovon K. Pramanik, Subrata Pramanik, Bimal K. Pramanik, M.K. Islam Molla and Md. Ekramul Hamid, "Hybrid Classification Algorithm for Knowledge Acquisition of Biomedical Data", International Journal of Advanced Science and Technology, Vol.44, July, 2012.

[4] N. Deepika and K. Chandra shekar, "Association rule for classification of Heart Attack Patients", International Journal of Advanced Engineering Science and Technologies, Vol. 11, No. 2, pp. 253 – 257, 2011.

[5]Mai Shouman, Tim Turner and Rob Stocker, "Integrating Decision Tree and K-Means Clustering with Different Initial Centroid Selection Methods in the Diagnosis of Heart Disease Patients", Proceedings of the International Conference on Data Mining, 2012 .

[6]K. Srinivas, B. Kavitha Rani and Dr. A. Govrdhan, “Application of Data Mining Techniques in Healthcare and Prediction of Heart Attacks”, International Journal on Computer Science and Engineering, Vol. 02, No. 02, pp. 250 - 255, 2011.

[7]Chaitrali S. Dangare and Sulabha S. Apte, “Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques”, International Journal of Computer Applications, Vol. 47, No. 10, pp. 0975 – 888, 2012.

[8]M A. Jabbar, Priti Chandra and B. L. Deekshatulu, “Cluster based association rule mining for heart attack prediction”, Journal of Theoretical and Applied Information Technology, Vol. 32, No.2, pp. 197 - 201, 2011.

[9]Olatubosun Olabode and Bola Titilayo Olabode, “Cardiovascular Accident Attack Classification Using Multilayer Feed Forward Artificial Neural Network with Back Propagation Error”, Journal of Computer Science, Vol. 8, No. 1, pp.18 - 25, 2012.

[10]M. Anbarasi, E. Anupriya and N.CH.S.N. Iyenga, “Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm”, International Journal of Engineering Science and Technology, Vol. 2, No. 10, pp. 5370 - 5376, 2010