Princess Nora Bint Abdulrahman
University College of Computer
and Information Science
Information Systems Department
IS 451D: graduation project 1,2

جـــامعـة الأميرة نـــــورة بنت
عبـــدالرحمن كليـــة علـــوم
الحـــاســـــب والمعلـومـــات
قســـم نظـــم المعلـومـــات
نـال 451د: مشـــروع تخرج١،٢

Effective Heart Disease Prediction System

EHDP System
For Early Risk Detection

A graduation project report submitted to college of computer sciences and information at PNU in partial fulfillment of the requirement for the degree of bachelor in Information Systems

CCIS, PNU
Riyadh, KSA
1441

Supervised By

Dr. Amal Al Rasheed

Prepared By

Nouf Alenzi	437003924
Ghadi Alamri	437003905
Sara Alhibss	437003947
Reem Alsalloum	437003837
Raghad Alshathri	436005142
Amjad Bin Salloom	434007341

Table of Contents

List of Tables

List of Figures

List of symbols and abbreviation

PNU	Princess Nora bint Abdul Rahman University
KSA	Kingdom of Saudi Arabia
DB	Database
SVM	Support vector machine
EHDP	Effective heart diseases prediction
CVD	Cardiovascular diseases
RIPPER	Repeated Incremental Pruning to Produce Error Reduction
REP	reduced error pruning
BN	Bayesian Network
MLPNN	Multilayer Perceptron Neural Network
BP	Backpropagation
LR	Logistic Regression
GUI	graphical user interface
WAC	Weighted Associative Classifier
SQL	style query language
CRISP-DM	Cross-industry standard process for data mining
HTML	Hypertext Markup Language
PHP	Personal Home Page
GPS	Global Positioning System
ASCVD	Atherosclerotic Cardiovascular Disease
MTBF	mean time between failures

Acknowledgment

The success and final outcome of this project required a lot of guidance and assistance from many people. We are extremely privileged to have got this all along the completion of our project. We are thankful for the assistance and help we got from faculty members at Information Systems department. Because of them, we were able to achieve our goal and get this project completed. Dr.Amal AL-Rasheed, our supervisor, was a great help to all of us. In spite of her busy schedule, she guided us step by step, with every little detail, that we needed. The project would not have been finished without her guidance. Also, with the support of our families, we were able to finish our task. Their encouraging words made us work harder. Finally, we would like to dedicate the efforts and outcomes of this project for all who suffer from heart disease in Saudi Arabia. We would like to tell them that we will be there for them in their fight against heart disease.

Abstract

Heart disease prediction systems are regarded as one of the most important subjects in the clinical data analysis sector. However, due to the complexity of heart diseases, the diagnosis depends on complex combination of clinical and pathological data which makes the diagnosing is challenging. Meanwhile, the healthcare industry has a significant amount of collected data that contains valuable information which could be mined to help with making effective decisions. The aim of this project was to propose a system that can detect heart diseases with maximum accuracy and predict the risk level of the disease by mining and analyzing datasets. We performed the classification experiment using Weka software. As a result of the experiment, we created a prediction model using support vector machine data mining technique to predict the occurrence of a heart disease. In odreder to build the model, the system used 14 medical features for prediction such as age, gender, fasting blood sugar, blood pressure, cholesterol, and obesity. Finally, we developed a website to help patients and doctors to diagnose heart disease.

Keywords

Disease Diagnosis, Data Mining, Machine Learning, Classification, Support Vector Machine, Principal Component Analysis.

Chapter 1

1. Introduction

Heart disease is a term that assigns to a large number of health care conditions related to the heart. These medical conditions describe the unexpected health conditions that directly control the heart and all its parts. Nowadays, heart diseases are one of the major causes of death globally [1]. Most people believe they are healthy because they are not showing symptoms but still could be at risk and have a high chance of devolving a heart disease.

Cardiovascular diseases (CVD) alone account for nearly half (46%) of all deaths in the kingdom of Saudi Arabia (KSA) [2]. Based on data from the National Health Accounts (NHA), the ministry of health in KSA has spent most of its budget on cardiovascular diseases which is one of the main issues that the government face [2]. Thus, providing good health care services is a challenge for the health care sector.

The health care sectors collect huge amounts of data that contain some hidden information, which is useful for making effective decisions [3]. In spite of being information-rich, the health care sector is still knowledge weak. There is a wealth of data possible within the health care systems. However, there is a lack of powerful analysis tools to identify hidden relationships and trends in data [4].

Cardiologists can generate a score that indicates a patient's heart disease risk level traditionally but it's not an exact prediction. Hence, the use of data mining methods could be useful in predicting heart diseases and turns the large collection of raw healthcare data into information that can help to make informed decisions and predictions [5]. The diagnosis of heart disease can offer an automated prediction about the heart condition of a patient so that further treatment can be made effective [6].

The aim of this project is to propose a system that can detect heart diseases with maximum accuracy and predict the risk level of the disease by mining and analyzing datasets. For providing appropriate results and making effective decisions on data, some advanced data mining techniques will be used. In this project, we will also develop a website to help patients and doctors to diagnose heart disease.

1.1 Problem Statement & Significance:

Heart diseases are one of the most life-threatening diseases that commonly infect adults and children, men and women all around the world. According to the Public Statistics Commission, the diagnostics of cardiovascular disease has unfortunately reached 1212 for every 100,000 cases of the population in Saudi Arabia (for a percentage of 1.21% of the population) (figure 1). Moreover, nearly half of all deaths in KSA are due to cardiovascular diseases.

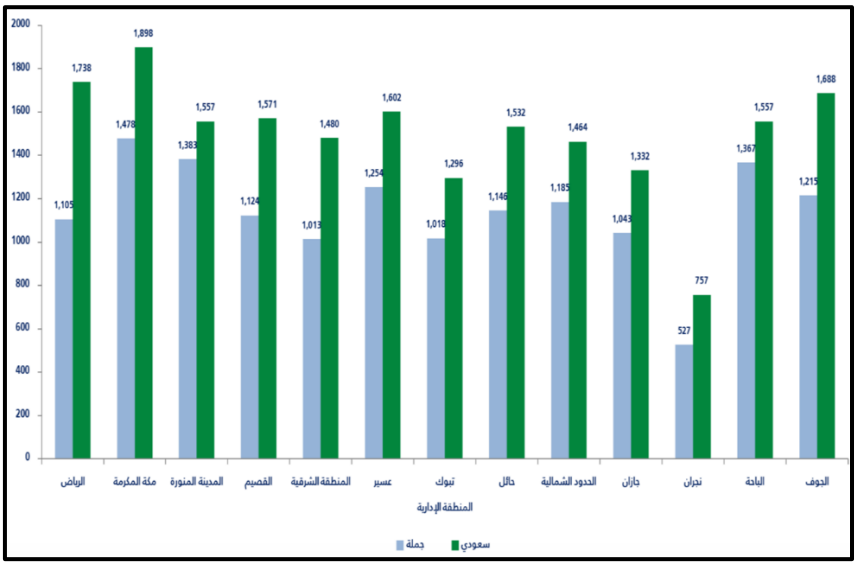

Figure 1 : Number of people that are diagnosed with heart diseases in different regions in Saudi Arabia [7]

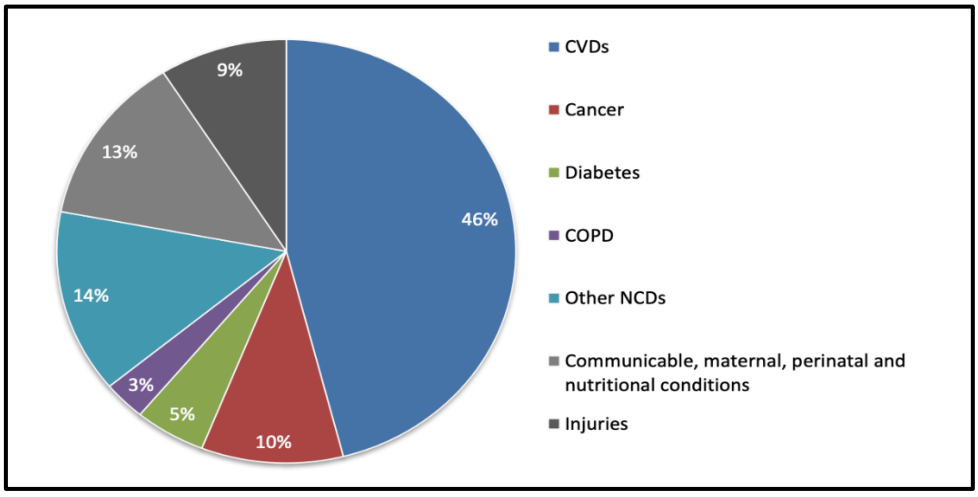

Figure 2: Percentage of different disease and the total deaths for all ages and both gender [2]

In fact, there are many different types of heart diseases generating various ambiguous symptoms making it difficult for doctors to diagnose efficiently. Meantime, health care industries collect a huge number of patients' data to be used in helping doctors diagnose diseases using traditional systems. These traditional systems are less effective than developing advanced systems that apply computer science methods/techniques to support doctors more affectively in predicting heart diseases and risks associated with it which makes the economic and social value of prediction very high. Thus, the system will add value to society by improving the mechanism of handling these cases in an advanced way using effective methods.

1.2 proposed solution:

The aim of this project is to develop a web-based system for heart disease prediction exploiting the capability of data mining methods. To achieve this aim, the objectives of this project are:

- To investigate the current work of data mining methods in heart disease prediction.
- To implement some advanced data mining methods and compare their accuracy.
- To build a model for heart disease prediction.
- To design and implement a web-based system that utilizes the built model for heart disease prediction.
- To provide some statistics and results.

1.3 Project Domain & Limitations:

We will develop a web-based system that could be used by patients and doctors to predict the presence of heart disease. The system will be used for heart disease and will not predict other diseases.

1.4 Definition of New Terms:

This report contains some new terms. These terms and their definitions are listed in table 1.

Table 1: Definition of new terms

Term	Description
Cardiovascular diseases (CVD)	It is a condition that involve the narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease.
Heart disease	Describe a range of conditions that affect the heart, heart disease includes blood vessel diseases such as coronary artery disease, heart rhythm problems, and heart defects you're born with (congenital heart defects).
Cardiologist	Cardiologists are doctors who are specialized in diagnosing and treating patients with heart diseases or blood vessels.
Prediction	A statement about what you think will happen in the future.
Data Mining	The practice of examining large pre-existing databases to generate new information.
Weka	It is a machine learning collection of algorithms for data mining tasks.
PyCharm	PyCharm is an Integrated Development Environment by JetBrains. It is used for development in Python and frameworks like Django.

Chapter 2

2. Background information and related work

2.1 Background Information:

2.1.1 Heart Disease

Heart disease describes a range of conditions that affects the heart. Diseases under the heart disease umbrella includes blood vessel diseases, such as coronary artery disease, heart rhythm problems (arrhythmias), heart attack, congestive heart failure, and heart defects you were born with (congenital heart defects). Each type of heart problems requires different treatment but may share similar warning signs [8].

World Health Organization has estimated that the morality rate caused by heart diseases will reach 23 million cases by 2030. This makes heart disease a major concern to be dealt with [7].

There is a study conducted among young Saudi women living in Al-Qassim, Saudi Arabia. The aim of the study was to assess the level of risk for cardiovascular diseases. The result denotes that only 15% of the sample was free of risk factors which means that young women in Al-Qassim have an unusually high risk for CVD. This result was limited to women living in Al-Qassim but what about women living in other districts in Saudi Arabia. Since the number of risk factors increases substantially between the ages of 20 and 35, there is a need to develop prevention programs to lower the CVD risk [9].

A wide range of industries including health care are already using data mining tools and techniques to take advantage of historical data. By using pattern recognition technologies and statistical and mathematical techniques to sift through warehoused information, data mining helps analysts recognize significant facts, relationships, trends, patterns, exceptions and anomalies that might otherwise go unnoticed [10].

2.1.2 Data Mining:

The information age and the computerization of almost every aspect of our lives have increased our capabilities to generate and collect data from various sources. Creating tremendous amounts of data done on a daily basis. This rapid growth in data collected and stored in data repositories has created the need for new automated tools and techniques to assist us in transforming the huge amounts of data into useful information and knowledge [11]. Data mining is a result of information technology which can be defined as the computer-assisted process of excavation and analyzing enormous sets of data and then extracting useful information from the data. Data mining tools predict behaviors and future trends, allowing various domains including health care institutions to make efficient, knowledge-driven decisions and answer questions that traditionally were time-consuming in order to resolve disclosing hidden patterns that help finding predictive information that experts may miss because it lies outside their expectations. [12].

2.1.3 Data mining techniques:

There are many different techniques in data mining using primary processes supervised learning and unsupervised learning as shown in (figure 3).

A. Supervised learning: is the data mining task of inferring a function using a labeled training dataset [13]. The main goal of supervised learning is prediction or classification. Common analytical models used in supervised data mining approaches are: Linear Regressions, Time Series, Classification or Regression Trees, Neural Networks and K-Nearest Neighbor.
B. Unsupervised learning: focus on understanding and describing the data to reveal hidden patterns within it. Unsupervised learning includes clustering, Association Analysis and Principal component analysis.

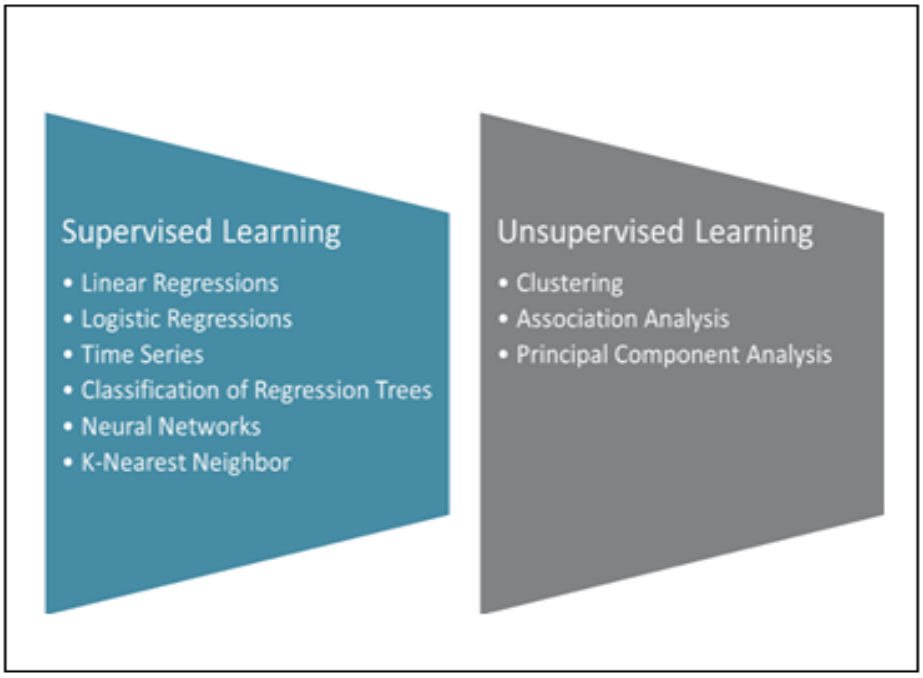

FIGURE 3: SUPERVISED AND UNSUPERVISED LEARNING [13]

In addition, there are some classification techniques that are very practical for heart diseases prediction, for instance:

A. RIPPER

RIPPER stands for Repeated Incremental Pruning Which Produce Error Reduction. This classification algorithm was proposed by William W Cohen. This algorithm is based on association rules with reduced error pruning (REP), it's a common and effective technique found in decision tree algorithms. In REP for rules algorithms, the training data is split into a growing set and a pruning set [14] as shown in figure 4:

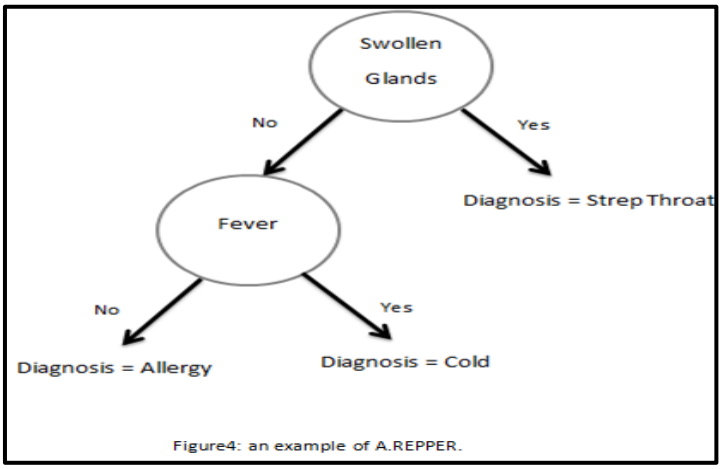

Figure 4: RIPPER algorithm [14]

B. Support Vector Machine

The SVM is a state-of-the-art maximum margin classification algorithm rooted in statistical learning theory. It's a classification method for both linear and non-linear data. Where non-linear mapping is used to transform the original training data into a higher dimension. Within the new dimension it searches for linear optimal separating hyperplane. With an appropriate nonlinear mapping to a sufficiently high dimension, in addition, hyperplane separate data from two classes. The SVM finds this hyperplane using support vectors and margins [15]. SVM performs classification tasks by maximizing the margin separating both classes while minimizing the classification errors [14].

C. Naïve Bayes Method

This classification method finds all records that are equal in predictor values by scanning the training dataset. Then the most prevalent class of the group is determined and assigned to the entire collection of observations. If a new observation's predictor variable equals the predictor variable of this group, the new observation will be assigned to this class. Due to the simplicity of this method a large number of records are required to obtain accuracy [14].

D. k-Nearest Neighbors Method

This classification method divides a training dataset into groups of k observations using a Euclidean Distance measure to determine the similarity between "neighbors". These classification groups are used to assign categories to each member of the validation training set [14].

E. Logistic Regression Method

A variant of ordinary regression which is used to predict the response variable, or the output variable, when the response variable is a dichotomous variable (a variable that takes only two values such as yes/no, success/failure, survive/die, etc.) [14].

F. Random forest

Random forest is a supervised learning algorithm. a big part of machine learning is classification, the fundamental concept behind random forest is a simple but powerful one. Random forest is useful with high dimensional data when we are dealing with subsets of data. It is faster to train than decision trees because we are working on a subset of features in this model, so we can work with many features
[14].

G. Neural Networks

The neural network is an analytical model inspired by the structure of the human brain, its neurons, and their connections. These models were originally created in 1940s but just recently gained popularity with statisticians and data scientists. Neural networks use inputs and, based on their magnitude, will “fire” or “not fire” its node based on its threshold requirement. This signal, or lack thereof, is then combined with other “fired” signals in the hidden layers of the network, where the process repeats itself until output is created [13].

2.1.4 Data mining tools:

PyCharm:
PyCharm is an IDE (Integrated Development Environment) by JetBrains. It is used for development in Python and frameworks like Django. Also, the developers can even customize the PyCharm UI according to their specific needs and preferences. Furthermore, they can extend the IDE by choosing over 50 plugins as well to meet their complex project requirements. PyCharm contains a lot of features and tools which makes it more popular.

Important Features:

- Code Editor.
- Code Navigation.
- Refactoring.
- Support for popular Web Technologies.
- Support for Python Scientific Libraries.

Important Tools:

- Database Tools.
- Visual Debugger.
- Built-in Terminal.
- Software Testing.
- Remote Development Capabilities [16].

WEKA:

WEKA is a collection of machine learning algorithms and data preprocessing tools designed to try out existing methods on new datasets in flexible ways. WEKA runs on almost any platform and It provides extensive support for the entire process of experimental data mining, including evaluating learning schemes statistically, preparing and visualizing the input data in addition to the result of learning.

WEKA provides implementations to a wide variety of learning algorithms that can be applied easily to datasets. It also includes a variety of tools for transforming datasets, such as the algorithms for discretization and sampling. with WEKA without writing any program code almost anyone can preprocess a dataset, feed it into a learning scheme, and analyze the resulting classifier and its performance. The WEKA workbench includes methods for the main data mining problems: clustering, regression, classification, association rule mining, and attribute selection. The ways of using WEKA are:

A. Apply a learning method to a dataset and analyze its output to learn more about the data.
B. Use learned models to generate predictions on new instances.
C. Implementing several different learners and compare their performance in order to choose one for prediction [17].

2.2 Related Work Survey:

Related work survey is very important and beneficial for any project, it will help leading the project to success. conducting the survey helped us to form a general idea and rollout any inefficient methodology or techniques. The table below summarizes the results of the survey.

Table 2: Related Work Survey

Title	Methodology	Results
Heart Disease Prediction Using Data Mining Techniques [18]	Random Forest. Decision trees. Naïve Bayes.	Random Forest provides perfect results compare to Decision tree and Naïve Bayes.
Effective Heart Disease Prediction System Using Data Mining Techniques [19]	Weka 3.6.11. Multilayer Perceptron Neural Network (MLPNN) with Backpropagation (BP).	System Created by Weka 3.6.11 tool predicts heart disease with 100% accuracy.
Identification of Significant Features and Data Mining Techniques in Predicting Heart Disease [20]	K-NN. Decision Tree. Naïve Bayes. Logistic Regression (LR). Support Vector Machine (SVM). Neural Network and Vote.	Experiment results achieved an accuracy of 87.4% in heart disease prediction.
Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques [21]	Neural Networks. Decision Tree. Naïve Bayes.	The results accuracy of the used techniques: 1. 100% for Neural Networks. 2. 99.62% for Decision Trees. 3. 90.74% for Naïve Bayes.
Backpropagation Neural Network for Prediction of Heart Disease [22]	Neural Networks. Decision Tree. Naïve Bayes.	good performance of the proposed algorithm compared to similar approaches.
Heart Disease Prediction Using Hybrid Genetic Fuzzy Model [23]	Genetic Algorithms Fuzzy logic.	The system provides an accuracy of 86%.
Human Heart Disease Prediction System using Data Mining Techniques [24]	Naïve Bayes. K-NN. Decision Tree Algorithm. Neural Network.	Adding a number of prominent attributes increases the accuracy of the prediction system.
Analysis of Coronary Heart Disease and Prediction of Heart Attack in Coal Mining Regions Using Data Mining Techniques [25]	Decision Trees. Naïve Bayes. Neural Network. Ruleset classifier. If -then. Neuro-fuzzy. Bayesian Network (BN). Support Vector Machine. Time Series Modeling.	Decision tree appears to be most effective for predicting the accuracy with 89%.
Predicting Heart Ailment in Patients with Varying Number of Features Using Data Mining Techniques [26]	Naïve Bayes. Random forest. Random tree. Clustering. Decision tree.	The prediction accuracy of heart ailment gradually increases irrespective of the classifier used to build the model.

2.3 Proposed & Similar System Comparison:

Our team constructed a comparison Table 3. It is a survey about similar programs and systems to our system, and how it is differentiated.

Table 3: Proposed & Similar System Comparison

	Proposed System	Similar System1 [27]	Similar System2 [28]	Similar System3 [29]	Similar System4 [30]
Problem solved	Predict heart disease using data mining techniques and big data.	Using data modeling and analysis tools to generate a knowledge-rich environment which can improve the quality of clinical decisions.	Design a GUI based Interface to enter the patient record and predict whether the patient is having heart disease or not by using the Weighted Association rule-based Classifier.	Predicting heart attack with a system that uses big data and data mining techniques	Evaluate the risk of being diagnosed with a cardiovascular disease using fuzzy tools and speculate risk levels for individuals.
Domain/Users	Patients & Doctor	Healthcare & Doctors	Patients & Doctors	Doctors, patients & medical students	Doctors
Design Methods	Using data mining classification techniques.	Use data mining classification modeling techniques, Decision Trees, Naïve Bayes and Neural Network.	Using Weighted Associative Classifier (WAC) as a Data mining technique to generate a rule base.	Naïve Bayes, Support Vector Machine, Simple Logistic Regression, Random Forest & Artificial Neural Network (ANN)	Fuzzy classification. Membership functions used for mapping.

Software & hardware	WEKA, HTML, PHP and Python.	Six phases of CRISP-DM methodology to build the data mining model, Data Mining Extension (DMX), SQL-style query language	Java Platform as front end, MS as backend, trained by using benchmark data from UCI machine learning repository.	Software: WEKA, TANAGRA, RapidMiner, MATLAB, Apache Mahout Hardware: Arduino/Genuino Uno, heartbeat sensor, electric buzzer, temperature and Humidity Sensor	PHP Language, Apache server used for development, HTML
Output	Website for heart disease prediction that visualizes results to the user and provides results and prediction based on data mining.	A web-based prototype heart disease decision support system that extracts hidden knowledge from a historical heart disease database	Intelligent Heart Disease Prediction System using Weighted Associative Classifiers that can be used in making an expert decision with maximum accuracy.	Web-based android application that uses efficient machine learning techniques to detect heart diseases.	Web-based Fuzzy Inference Tool for cardiovascular risk assessment that allows the health professionals to provide a prediction of heart disease risk, it can also predict other types of risks
Features	1. Patients create profile and login to the system. 2. Use graphs and visualization to help the patient	1. Display the results both in tabular and graphical forms. 2. Two models were used to test the effectiveness : lift chart	1. Implemented in remote areas 2. imitate human diagnostic expertise for treatments	1. patients are allowed to input the parameters of heart disease from anywhere on the application interface	1. the system allows the user to add variables such as age, gender, height, weight, blood pressure. 2. Use inputs such as Gender, Blood pressure,

	understand the results. 3. Track the symptoms and signs 4. Record the medication and add prescriptions. 5. Use GPS to display the nearest hospital/health center to the user. 6. Offer information's about cardiovascular doctors to the user and the doctor's background. 7. High accuracy of prediction. 8. Use inputs such as Gender, Age, Blood pressure and other	and classification matrix 3. provide effective treatment at a lower cost.	3. can be updated when the new training dataset is available 4. user-friendly 5. web-based	2. view the level of getting heart disease 3. patients can communicate with the doctors through video call 4. doctors related to heart disease can be found by searching their phone number 5. can also be used for other diseases alongside heart disease	Medication and knowledge 3. Use a maximum degree of membership

	related attributes.				
Limitations	The system is only available in Saudi Arabia.	1. The drill-through feature to access detailed patients' profiles is only available in Decision Trees. 2. The relationship between attributes provided by Neural Network is more difficult to understand.	The system only uses dataset which are used for research purpose from UCI machine learning.	1. No algorithm reached an accuracy level of more than 90 percent in heart disease prediction using the same number and features. 2. When feature numbers reduce, the performance measure also will be reduced.	With varying degrees of membership, users may face different risk ranges.

To summaries, all systems are designed to solve the same problem and predict heart diseases using similar methodologies and features. The system targets similar domain users mostly doctors and patients. All the systems reviewed in the table used different softwares such as WEKA, HTML, PHP and Machine learning.

Chapter 3

3 System Analysis

3.1 Requirements Specification

3.1.1 Requirements Gathering

The process of requirements gathering is an important and essential part of any research study. Data collection helps us to fully understand all of the system aspects and it is critical to the system in order to succeed. Also, It's crucial for the analyst to gather suitable, correct complete data from a reliable source to give value to the system and avoid failure. We used robust requirements gathering approach in addition to key techniques and methods, such as:

A. Brainstorming

Brainstorming is a sufficient method that helps us to generate creative ideas and find solutions to any problem that might face us. As a group we had many effective meetings and brainstorming sessions discussing in detail heart diseases, the different ways of diagnosing and what are the system functions that we could give to the users. As an outcome with the help of our supervisor, we decided to use the heart disease datasets in UCI. Also, another result of the sessions was agreeing on using python programming language for building the model, WEKA tool for testing and visualizing also HTML language and PHP for building the website.

B. Literature review and similar tools analysis

A good literature review is a necessity for any academic research. It helps us with evaluation, comparison and classification. In our analysis process we reviewed dozens of Journals and articles in addition to testing various similar systems in the form of a website or a phone application. As a result, after reviewing other researchers work, we formed a general idea of how to build our system and avoid inefficient prediction methods and techniques. At last, analyzing those papers will help us design a unique system using a new algorithm.

C. Interviews

Interviews of potential users and understanding their perspectives are important in creating an efficient system. By knowing the user's expectations of the system, we will be able to meet those expectations in addition to the user's needs. One-on-one interviews are one of the primary sources of requirements gathering therefore we decided to interview and collect information from several Saudi cardiologists.

- Interviews Summary:

The users of the website are mainly doctors and patients and that's why we interviewed three Saudi doctors, and who are:

A. Dr. Abdulrahman Al-Qahtani, assistant professor and consultant cardiologist, interventional radiology (catheter) and replacement of heart valves without surgery. Laurentian University Canada.
B. Dr. Walid Al-Harbi Consultant Cardiac and Catheterization at King Saud University specializes in the complex catheterization of arteries and valves and peripheral arteries. The head of the cardiac catheterization at the King Fahd Center for Cardiology.
C. Dr. Mohammed Kurdi Consultant Cardiology and Catheterization and the President of the Saudi Society for Interventional Cardiac Catheterization.

We gave the doctors a brief about the system and we asked them the following questions:

> We are students from Princes Noura bint Abdulrahman University, Faculty of Computer Science and information specializing information systems.
>
> We are currently working on our graduation project which aims to provide and develop an effective system for predicting heart disease, based on data from patients from Saudi hospitals and the probability of developing the disease and the level of risk.
>
> The project aims to help doctors diagnose diseases with high accuracy and effectiveness, which also helps patients track their health and the likelihood of their illness.
>
> We hope you can help us answer some questions to help us collect the necessary data
>
> 1. What are the most important symptoms of heart disease?
> 2. What are the factors that promote the injury?
> 3. In your opinion, what are the means that facilitate communication between the patient and the cardiologist?
> 4. Are there systems used by doctors to help predict heart disease and determine its severity?
> 5. What are the services you want the system to provide?
> 6. Do you have an interest in using the system?

All doctors agreed that the most important symptoms are Chest pain, palpitations, chest pressure, shortness of breath, fainting and swelling in the lower limbs and the Factors that promote the injury are smoking, High blood pressure, High diabetes, Cholesterol, Family history and Obesity. Regarding ways of communication with the patients, doctors said that having an account on social media, talking with patients by phone or via email and review through video calls are all means that facilitate communication with the patients. Moreover, doctors have given us systems that they use to help them to predict and diagnose diseases such as ASCVD Risk Estimator [31], Grayscore [32] and MDCalc [33]. doctors want the system to provide probabilities and associate the risk of the disease. They also want the website to be user-friendly and easy to use. All doctors have a high interest in using the system. In addition, we interviewed people who have some symptoms or have a family history with heart diseases. We asked if they are interested in using our system to predict the probability of developing heart disease. Their answers were yes it will be helpful.

3.2 Requirements Analysis

3.2.1 Software/Hardware Requirements

A. Hardware

- Laptop with the following specification.
- RAM: Recommended 8GB-256GB.
- CPU: Recommended i4 core -i9 core.
- Operating system: Windows 10.
- Hard disk space: 4 GB of available disk space.

B. Software

- PyCharm and Anaconda for building the prediction model.
- WEKA tool for testing.
- Wix for designing interfaces.
- Notepad++ using HTML and PHP for creating the website.

3.2.2 Non-Functional Requirements

A. Security /privacy: The system should only provide access to authorized users and protect sensitive information.

B. Availability: The system will be available for users at all times.

C. Usability: The system will provide a simple easy to navigate interface.

D. Reliability: The system should provide a long MTBF (mean time between failures).

E. System quality: The system features have a high set of standards.

F. Accuracy: The system data are 100% accurate and precise.

3.2.3 Functional Requirements

A. **System and Admin Functional Requirements**

- The system will make the process of prediction.
- The Admin will verify and manage login information and test data information.
- The system will link the user with the nearest hospital/ health center.
- The system will provide educational health information to increase awareness.

B. **Patient Functional Requirements**

- The patient will be able to register an account with valid information.

- The user will be asked by the system if he/she is a doctor or a patient.
- The patient will be asked to enter the symptoms and signs.
- The patients will be able to view the nearest hospital based on their location.

C. **Doctor Functional Requirements**

- The doctor will be able to register an account with valid information.
- The user will be asked by the system if he/she is a doctor or a patient.
- The doctor will be asked to enter the symptoms and signs.

3.2.4 Use case diagram:

The Use case diagram is a graphic depiction of the interactions among the elements of a system. As shown in figure 5.

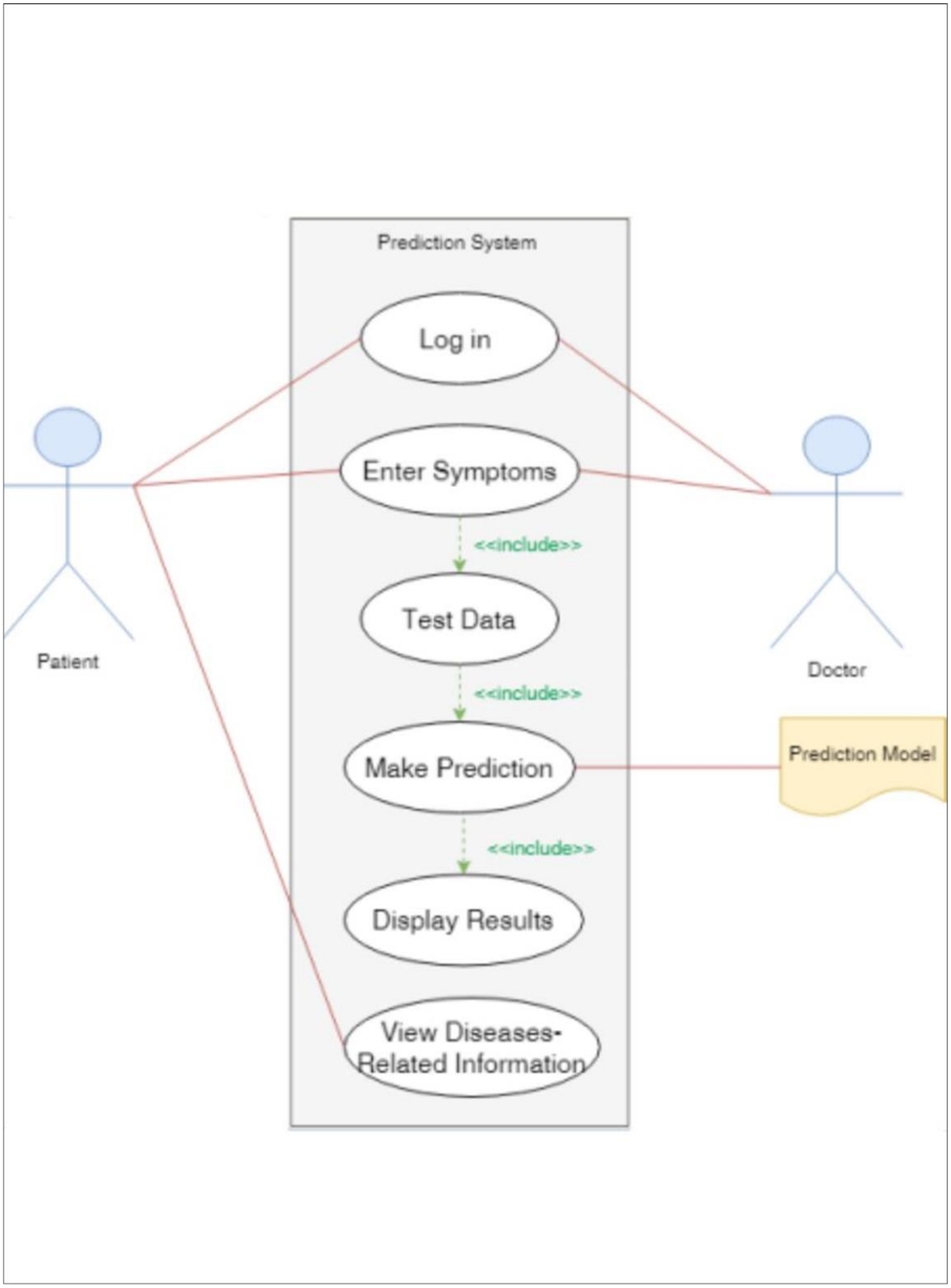

Figure 5: EHDP system website Use-case

Use Case Description

Table 4 : Log in use-case

Use-Case Number	UC_1
Use-Case Name	Log in
Actor	Admin, Doctor and patient
Description	Enable the user to gain access to the website by entering valid data.
Pre-Condition	None
Post-Condition	The user will be granted access to the website and the home page will appear.
Basic Flow of events	A. The user will access the website. B. The user will enter email and password then press the login button. C. The system will verify login information. D. if the information is invalid the user won't be granted access. E. if the user enters valid information, he will be granted access.

Table 5: Enter Symptoms Use-Case

Use-Case Number	UC_02
Use-Case Name	Enter Symptoms
Actor	Patient, Doctor
Description	The User will enter the symptoms to the prediction model.
Pre-Condition	Access to the system by Logging in.
Post-Condition	Data will be tested by the system.
Basic Flow of events	A. The user will access the make prediction page and fill the form with accurate data. B. The user will press the predict button.

Table 6: Test data use-case

Use-Case Number	UC_03
Use-Case Name	Test Data
Actor	Patient, Doctor
Description	After the user enters the symptoms, the prediction model will test the symptoms.
Pre-Condition	Enter symptoms.
Post-Condition	The system will review the result to the user.
Basic Flow of Events	A. After the use entered the Symptoms. B. The prediction model will test the data as part of the prediction process.

Table 7: Make prediction use-case

Use-Case Number	UC_04
Use-Case Name	Make Prediction
Actor	Patient, Doctor
Description	Predict if the user has a heart disease based on the data entered previously.
Pre-Condition	Data must be entered by the user and tested.
Post-Condition	Prediction is made and the result is shown to the user
Basic Flow of events	A. The prediction model will complete the prediction process. B. The prediction model will predict the outcome. C. Provide the last outcome.

Table 8: Display result use-case

Use-Case Number	UC_05
Use-Case Name	View Result
Actor	Patient, Doctor
Description	Enable the user to view results.
Pre-Condition	Make prediction.
Post-Condition	The website displays the result to the user
Basic Flow	A. The system will display the result of the prediction model to the user.

Table 9: View Disease-Related Information use-case

Use-Case Number	UC_06
Use-Case Name	View Disease-Related Information
Actor	Patient
Description	Enable the Patient to View Nearest Hospitals and View Educational Health Information to Learn more about Their Disease, How To control it and Provide Advises and Statistics to Improve Awareness
Pre-Condition	Access to the system by Logging in.
Post-Condition	The Website will View the needed Information
Flow of event	A. The Patient should login to the website. B. The system links the Patient with the nearest hospitals based on his/her location. C. The Patient will access "for your health page" to view Educational Health Information to Learn more about Their Disease.

3.2.5 Data flow diagram (DFD):

The Data flow diagram is a way of representing a flow of data of a process or a system. The data flow diagram Provides information about a process itself and the outputs and inputs of each entity as shown in figure 6.

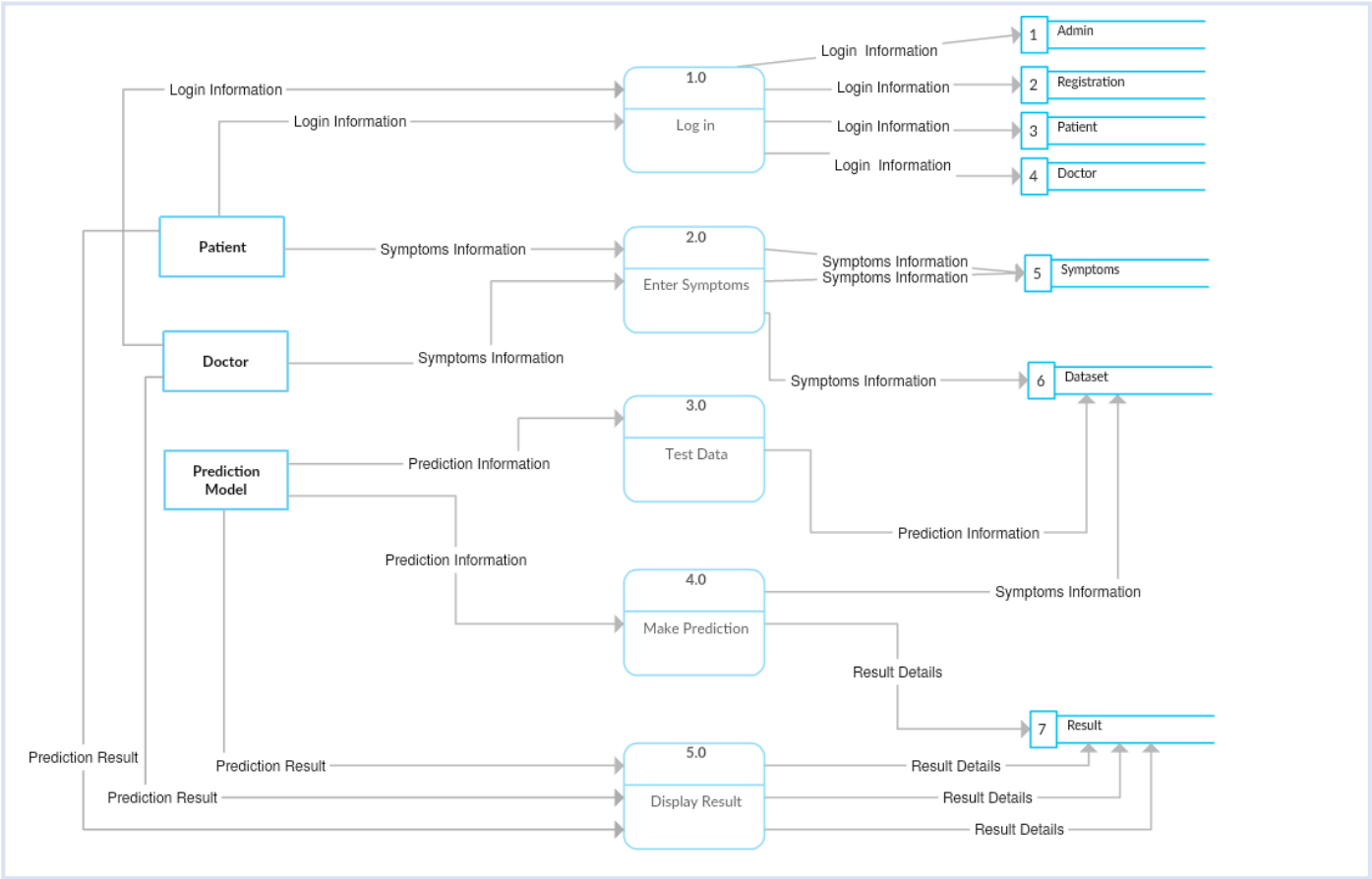

Figure 6: Data Flow Diagram

3.2.6 Entity Relationship Diagram (ERD):

Entity Relationship diagram is a graphical representation of entities and their relations to each other stored in a database as shown in figure 7.

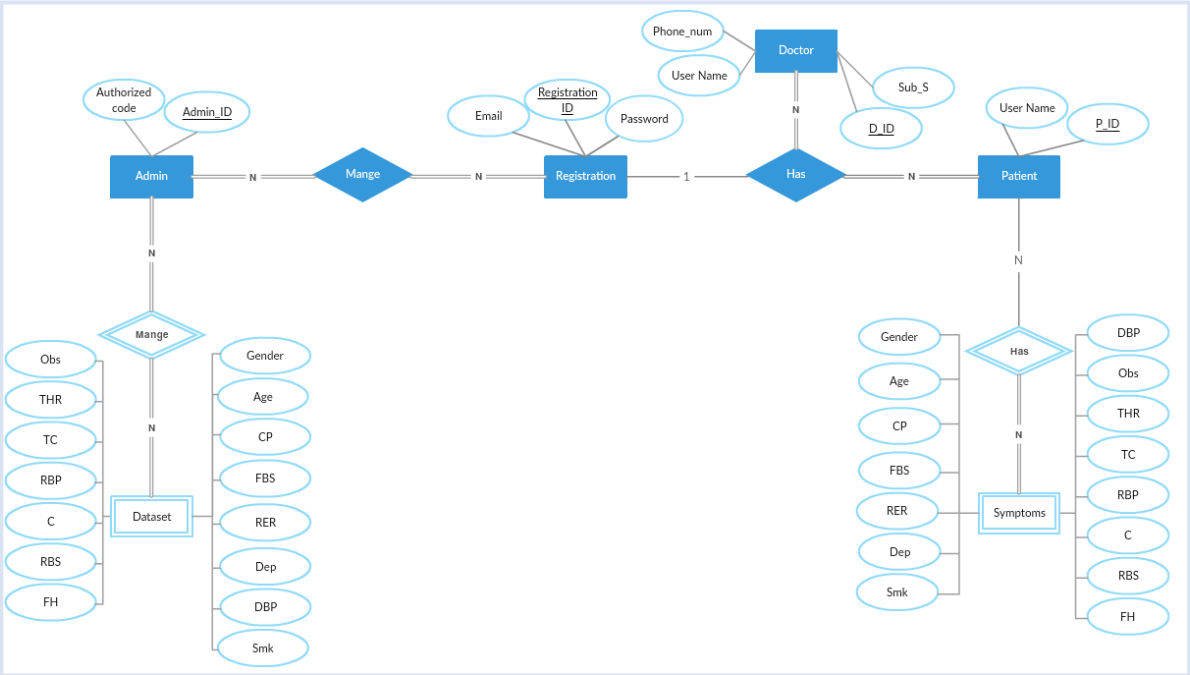

Figure 7: Entity Relationship Diagram

Chapter 4

4.System Design

4.1 System Architecture:

The System architecture provides the conceptual model that defines the structure. There are three components that make up the system as shown in Figure 8.

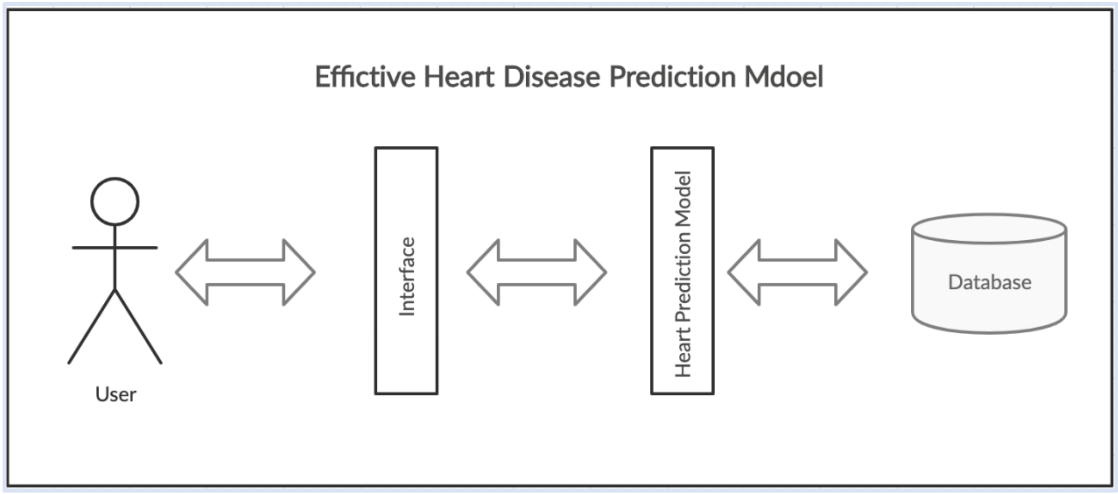

Figure 8: System Architecture

Interface:

This component provides services for the user. It takes the patient's data as input and the prediction result as output.

EHDP prediction model:

This layer contains the prediction model that we will use to investigate the relationship between future or unknown outcomes and baseline health states among people with specific conditions. The prediction model should predict accurately and be relatively easy to use. If a prediction model provides an inaccurate prediction of future occurrences It will be misleading. On the other hand, if the model has high predictability power but is difficult to apply it will not be commonly used.

Steps to develop our prediction model:

A. Preparation for establishing EHDP model:

The aim is to develop an accurate and useful prediction model with multiple variables using comprehensive datasets. Firstly, we choose the proper datasets for the model. The category of target users will determine the selection and handling process of multiple variables, which will affect the structure of the prediction model.

B. Dataset selection:

The dataset is one of the most important components of the prediction model—often not under investigators' control—and ultimately determines its quality and credibility; however, there are no rules for assessing the quality of data. Yet, there is no such thing as perfect data and perfect model. It would be reasonable to search for a best-suited dataset.

C. Handling dataset Variables:

The dataset must be evaluated to extract the most important and needed variables in the prediction model. Then the dataset must be cleaned to get rid of undesirable variables and to handle missing data. Missing data can occur various reasons, including uncollected (e.g., by design), not available or not applicable, refusal by respondent, dropout, or "don't know." To handle this issue, we may consider the imputation technique, dichotomizing the answer into yes versus others, or allow "unknown" as a separate category.

D. Model generation:

We will build a prediction model that uses data mining classification techniques. that can infer characteristics of predicted class from a combination of other data. The task of data mining is to build models for prediction of the class based on selected attributes. We apply the following algorithms, RIPPER, Support Vector Machine, Naïve Bayes Method, k-Nearest Neighbors Method, Logistic Regression Method, Discriminant Analysis Method, Neural Networks and Decision Tree algorithms.

E. Model evaluation and Validation:

After generating the model, we should evaluate the predictive power of our proposed model using an independent dataset, where truly an external dataset is preferred whenever available.

Database:

This layer contains the available database which contains the data related to doctor and patent profile information and the dataset used to build the model.

4.2 User interface design:

In this section, we introduce the design of our website interfaces.

4.2.1 Home page:

The home page of our website contains a header with three buttons that shows the services that will be provided by our website to the users either Make Prediction or More as shown in *Figure 10*. In addition, the homepage body contains a Login form. If the users don't not have an account, they can register by clicking the " register here " hyperlink as shown in *Figure 9*. The Home page also contains "About EHDP system", which displays information about the service provided and the target users.

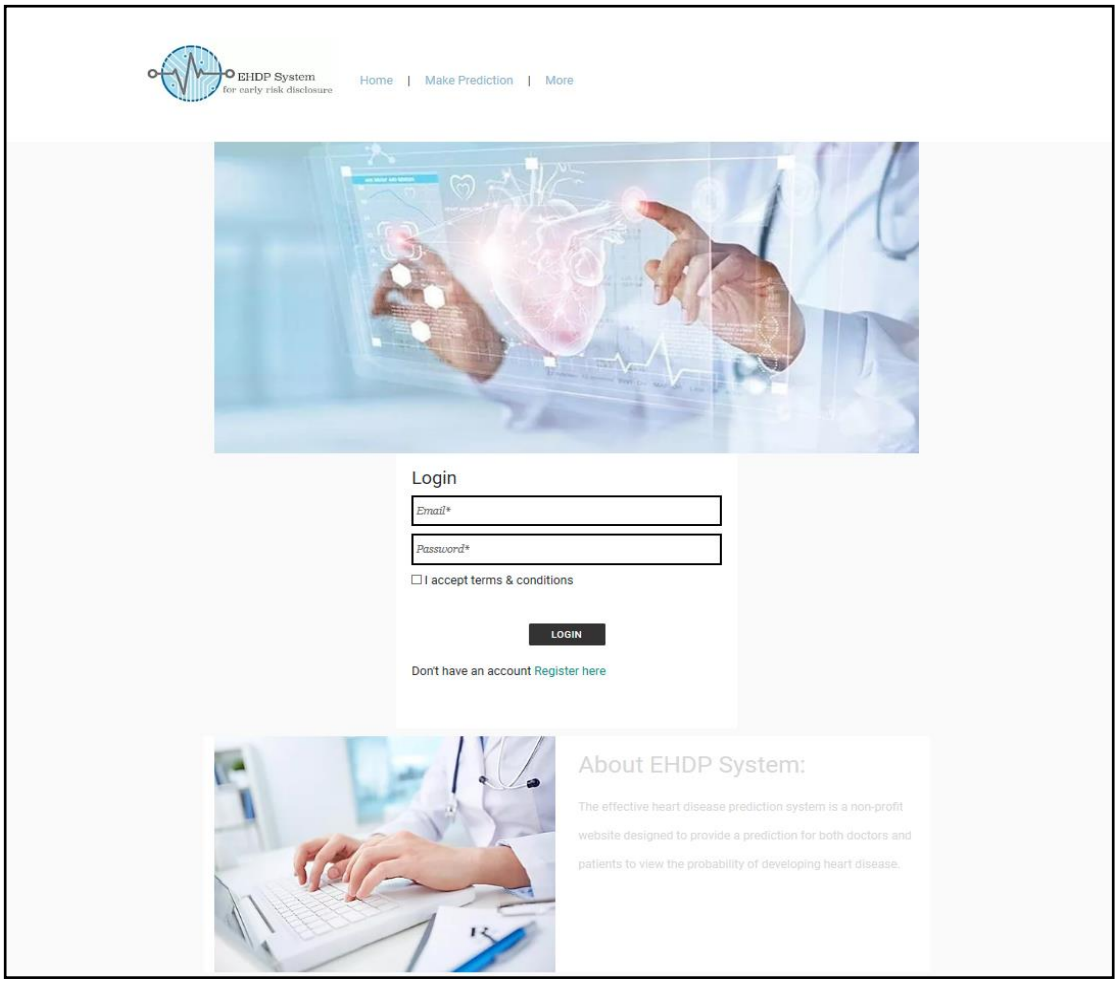

Figure 9: Website home page

4.2.2 More page:

The More drop list shows the other services in the website.

Figure 10: More (drop LIST)

4.2.3 Registration Page:

This page allows the user to create an account by filling the registration form. This page has a hyperlink to return to the home page.

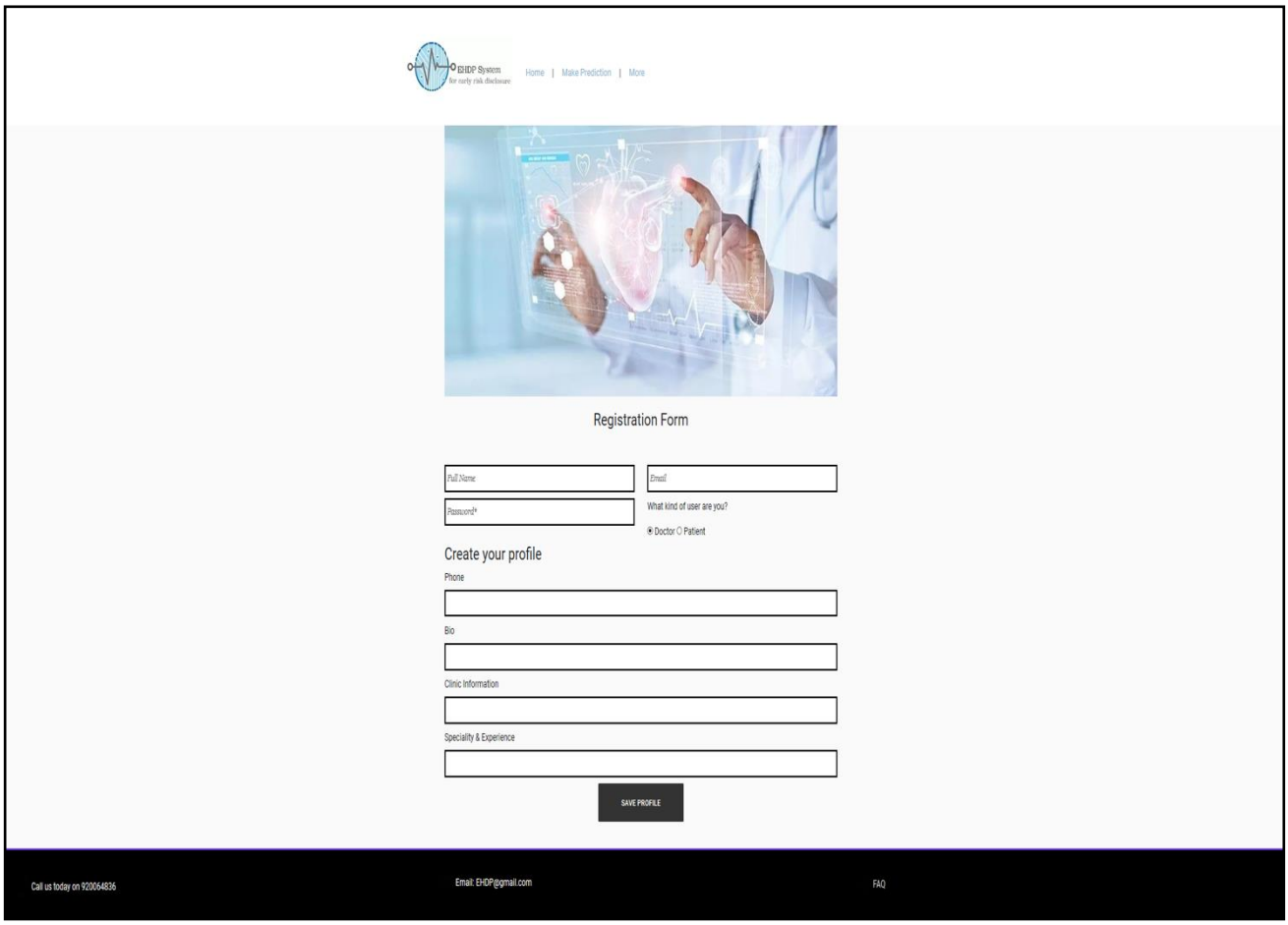

Figure 11: Registration Page.

4.2.4 Make Prediction Page:

Once the user clicks the "Make Prediction page", the Add Data form will be shown. Therefore, the user will enter the required data in the form and click predict to get the result. This page has a hyperlink to the Result page as shown in *Figure 13.*

Figure 12: Make Prediction Page

4.2.5 Result Page:

After clicking the predict button, the system will display the result to the user.

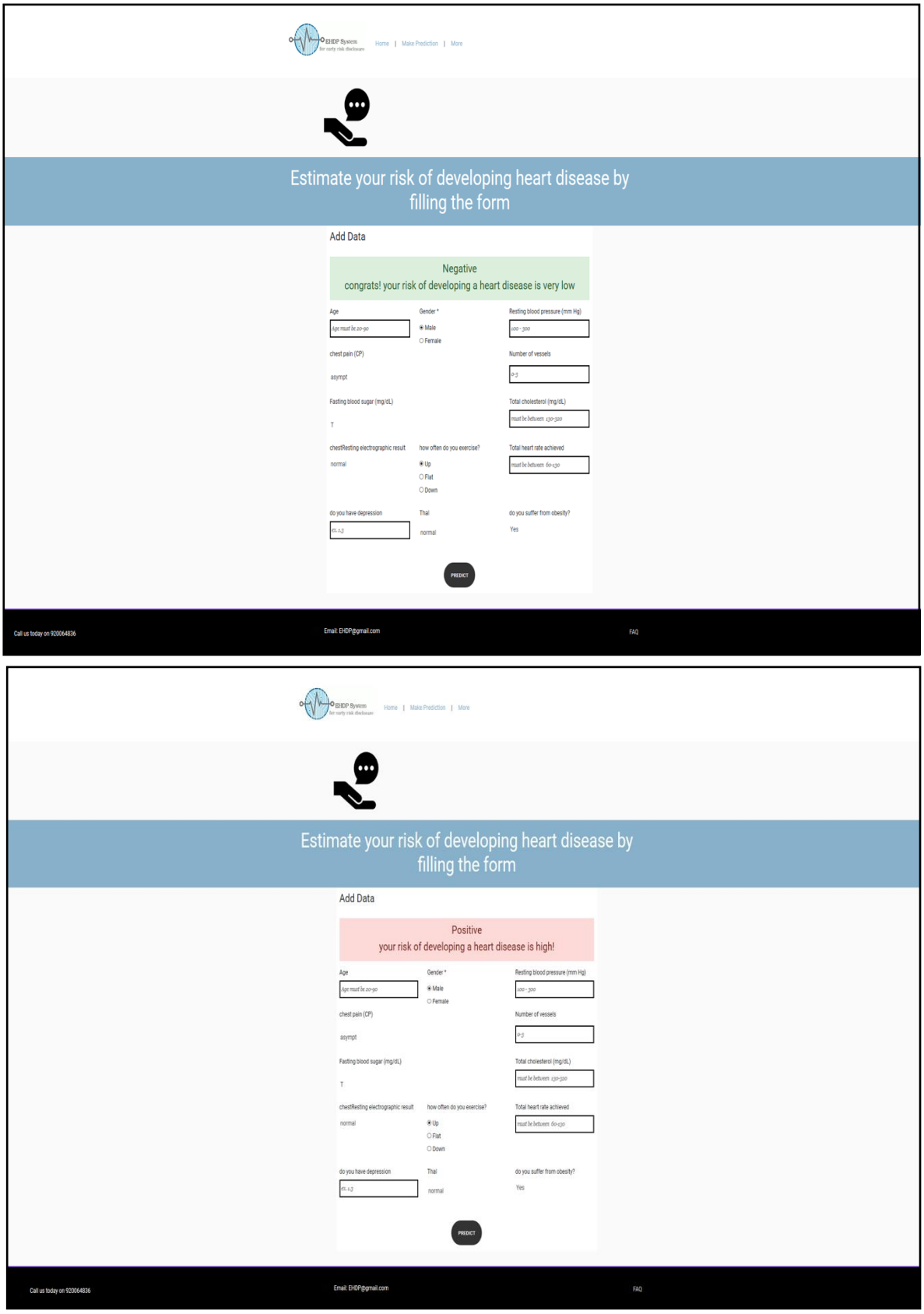

Figure 13: Result Page.

4.2.6 Tips page:

This page contains tips directed to heart disease patients to help them better manage their condition.

Figure 14: Helpful Tips for Patients Page.

4.2.7 Facts page:

This page provides statistics and facts about various heart diseases to increase awareness.

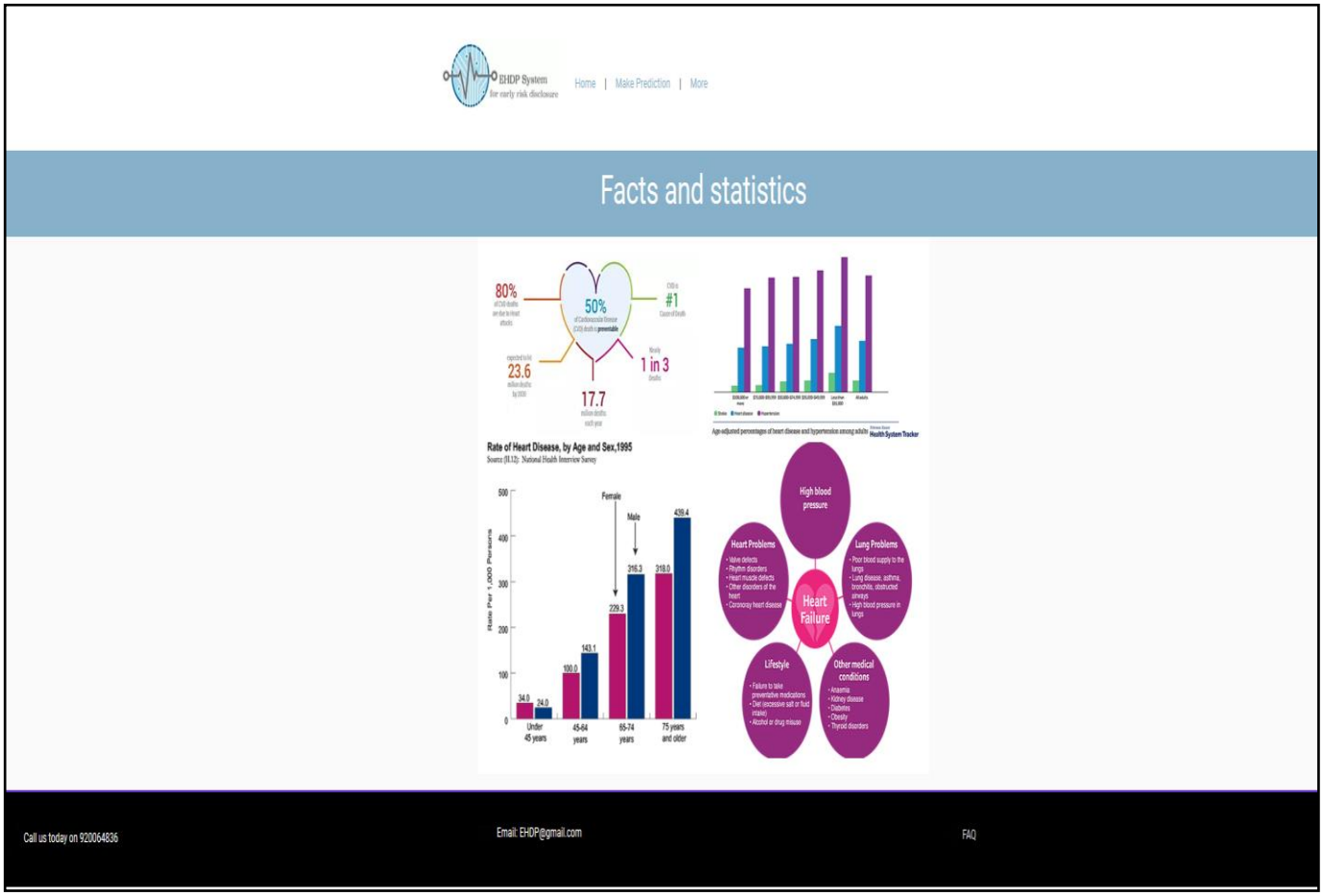

Figure 15: Facts page

4.2.8 View Nearest Hospital Page:

This page will display a map to show the users the nearest hospitals based on their location.

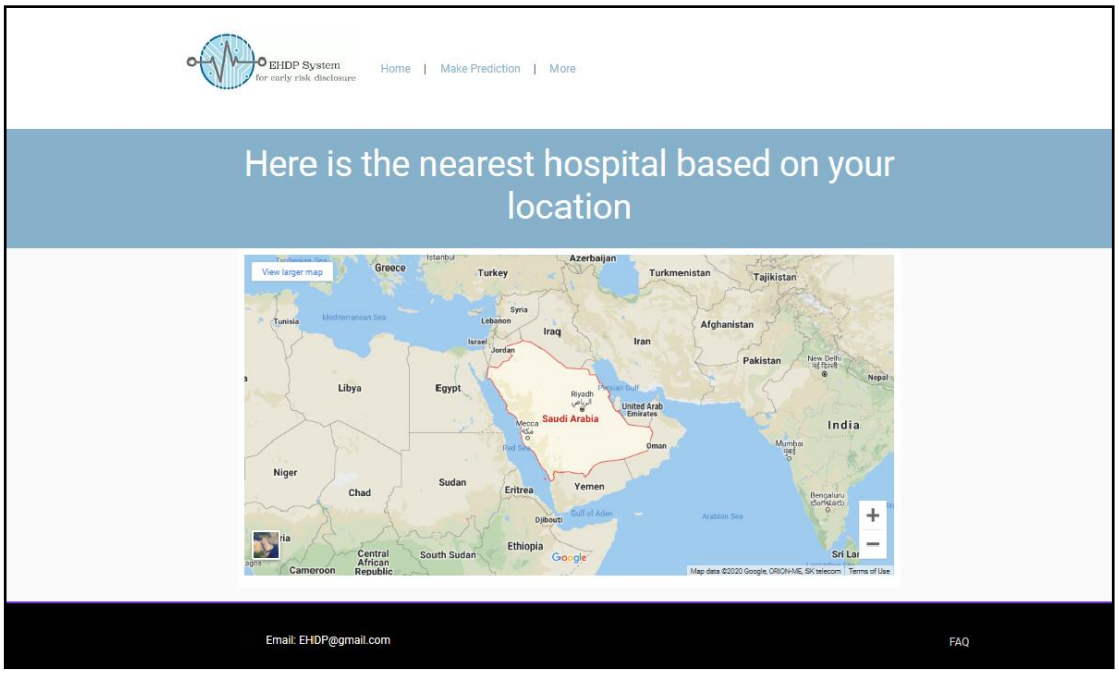

Figure 16: View Nearest Hospitals page

4.2.9 FAQ page:

This page answers the users common questions regarding the website and the system.

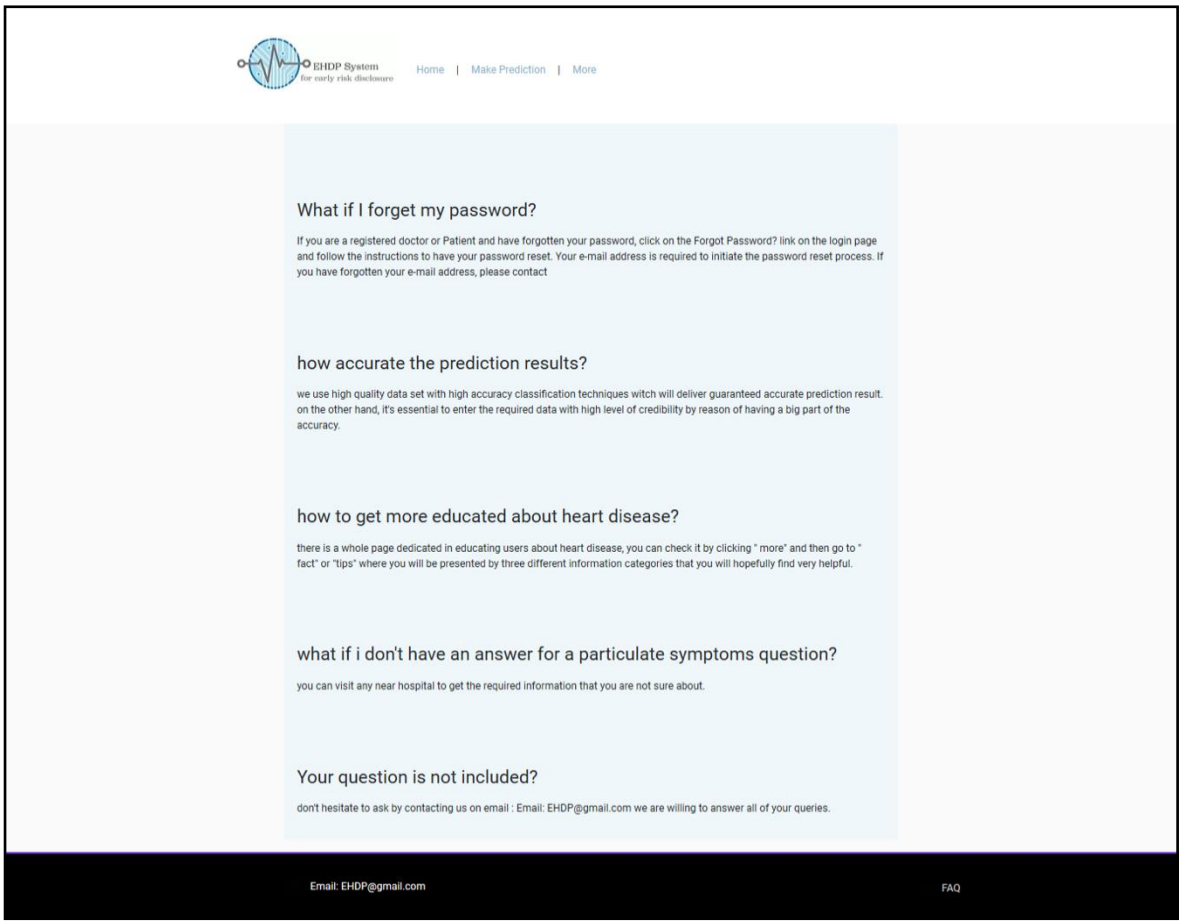

Figure 17: FAQ page.

Chapter 5

5. Implementation

In this chapter, we will describe the implementation phase in detail, starting from implementation requirements going through all implementation processes, ending with the final system. First, we will describe the methodology that we followed to implement our system. The methodology consists of six steps: get data (heart disease dataset), pre- processing, feature selection, classification, building the prediction model and evaluation. Moreover, we will describe the implementation of the website and the integration with the model.

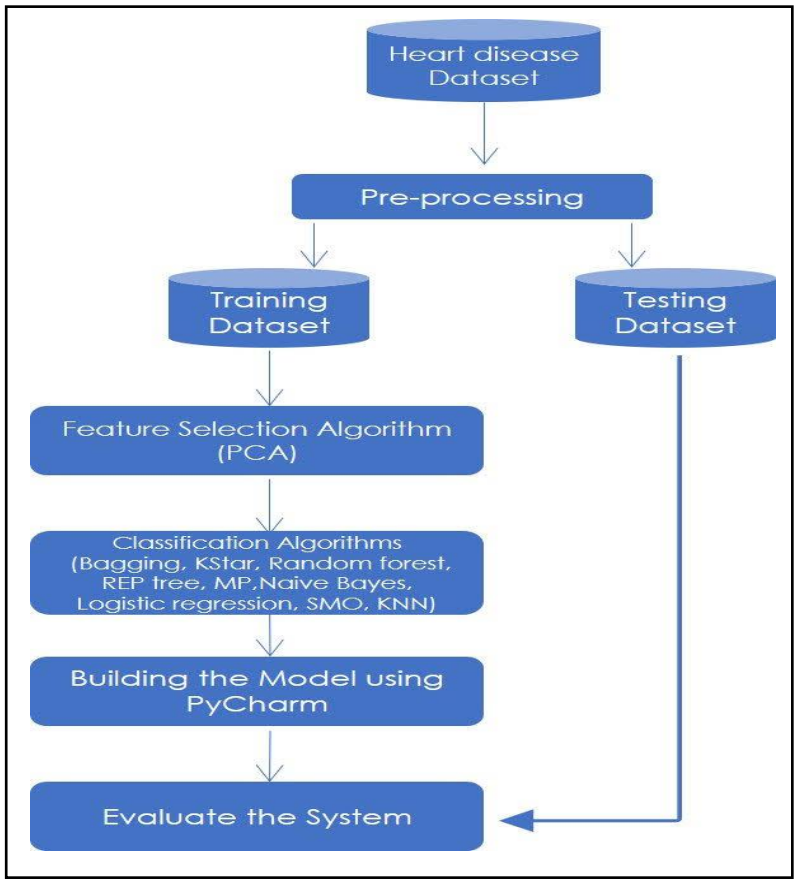

Figure18: Methodology.

5.1 Implementation Requirements

In this part, we will provide the implementation requirements including software and hardware used.

A. Hardware requirements:

Laptop with the following specifications:

- RAM recommended 8GB-256GB.
- CPU recommended i4 core -i9 core.
- Operating System windows 10.
- Hard Disk Space 4 GB of available disk space.

B. Software requirements:

- *PyCharm:* It is the integrated development environment used in computer programming, specifically for the Python language [34].
- *Html and Php:* It is programming languages to implement the website.
- *WEKA:* It is an open-source software that provides an implementation of several machine learning algorithms and tools for data preprocessing [35].
- *Anaconda***:** It is a private Python distribution that is available free and open-source with the main goal of facilitating the management of specific Python libraries [37].
- *Xampp*: It is an open-source program, we used it to run the website and manage the database on localhost [38].
- *TeamViewer*: It is a software package for controlling a remote computer, creating video meetings, as well as a file transfer service between devices [39].

5.2 Implementation Details.

In this part we will describe the implementation process in detials.

5.2.1 Get data:

First, the data that we used in our project is Cleveland Heart Disease dataset which is found in the UCI machine learning repository, an online open-source repository that is used by the machine learning community for the empirical analysis of machine learning algorithms. The dataset consists of 303 individual's data and 14 attributes, 8 nominal and 6 numerical attributes. The visualization of all attributes is shown in *figure 18*. The reason behind choosing this dataset is that the Cleveland Heart Disease dataset is cleaned, well defined, and the quality of its features has been proven through the usage by the research community. Also, the 14 attributes of the dataset are critical characteristics of heart diseases playing a major role in diagnosing heart diseases which fulfill the purpose of our project to predict the occurrence of a heart disease. The dataset uses a large range of instances for each attribute which increases the accuracy of the dataset and thus the prediction. [40].

The dataset attributes are:

A. Age.
B. Sex.
C. Chest-pain type which displays the type of chest-pain experienced.
D. Resting Blood Pressure.
E. Cholesterol.
F. Fasting Blood Sugar.
G. Resting ECG which displays resting electrocardiographic results.
H. Max heart rate achieved.
I. Exercise induced angina.
J. ST depression induced by exercise relative to rest.
K. Slope of the peak exercise ST segment.
L. Colorful by fluoroscopy.
M. Thal which displays the results from a thallium heart scan.
N. Diagnosis of heart disease and it displays if the individual is suffering from heart disease.

The detail attributes information is shown below:

Table 10: Attributes of Heart Disease Dataset

Number	Attributes	Discribtion	Type	Domain
1	Age	Patient age in years	Numerical	29 to 77
2	Sex	Gender	Nominal	1 = female 2= male
3	Cp	Chest pain type	Nominal	1: typical angina. 2: asymptomatic. 3: non-angina pain. 4: atypical angina.
4	Trestbps	Resting blood pressure	Numerical	94 to 200
5	Chol	Serum cholesterol	Numerical	126 to 564
6	Fbs	fasting blood sugar > 120 mg/dl	Nominal	1 = true 0 = false
7	Restecg	Resting electrocardiographic results	Nominal	1: left ventricular hypertrophy . 2: normal. 3: having ST-T wav abnormality.
8	Thalach	Maximum heart rate achieved	Numerical	71 to 202
9	Exang	Exercise induced angina	Nominal	1 = no 2= yes
10	Opk	Oldpeak	Numerical	Continuous (0 to 6.2)
11	Slope	The slope of the peak exercise ST segment	Nominal	1: up sloping. 2: flat. 3: down sloping.
12	Ca	Number of major vessels	Numerical	0 to 3
13	Thal	Defect type	Nominal	1 : fixed defect 2 : normal 3 : reversible defect.
14	Num	Diagnosis of heart disease	Nominal	0: < 50% diameter narrowing. 1: > 50% diameter narrowing.

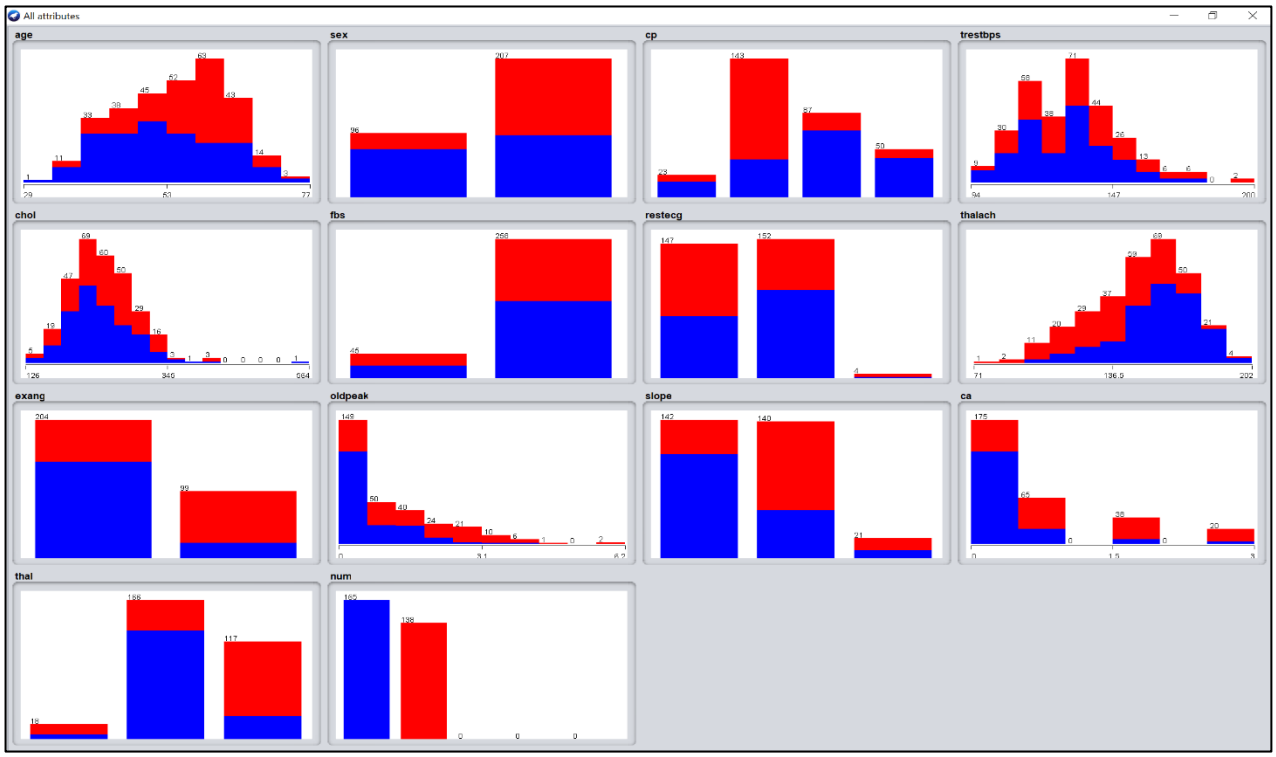

Figure 18:Attributes visualization

5.2.2 Preprocessing:

This stage is where cleaning, normalizing and balancing data are done. The data used in our project is public dataset from UCI. Therefore, the data was already cleaned, normalized and processed. The classification experiment was carried out under Weka environment. Due to the small number of selected features, 10-fold cross-validation was used which divides the dataset to training and testing datasets. We applied K-cross validation to train then test our models. K-Cross validation divides the dataset D into K equal-sized parts, called folds, namely D1, D2, ..., DK. Each fold Di is, in turn, treated as the testing set, with the remaining folds comprising the training set D\Di=Sj6=I Dj. After training the model Mi on D \Di, we assess its performance on the testing set Di to obtain the i-th estimate, usually, K is chosen to be 5 or 10.

For the purpose of avoiding instable operation results, each experiment runs 10 times, and the optimal classification accuracy was selected for comparison between each method used. After conducting the cross-validation, we evaluated the effectiveness of all classifiers in terms of time to build the model, correctly classified instances, incorrectly classified instances as shown in *table 1*.

5.2.2.1 Feature Selection (PCA):

In an attempt to improve the performance of the prediction model, we applied features selection algorithms. We applied principal component analysis (PCA) feature selection. PCA was invented in 1901 by Karl Pearson. It is "a technique used to emphasize variation and capture strong patterns in a dataset." it is used for identification of a smaller number of uncorrelated variables known as principal components from a larger set of data. It has helped data to be easier to visualize and explore [41]. The main goal of the principal component analysis is to reduce the feature of a dataset called dimensionality that contains many variables that are related to each other while maintaining and retaining the variation and spread in the dataset [42]. Figure 19 displays the dataset after applying PCA. After testing each method with PCA, we compared the accuracy of the classifiers before and after applying PCA feature selection as shown in *table 10* and *table 11*.

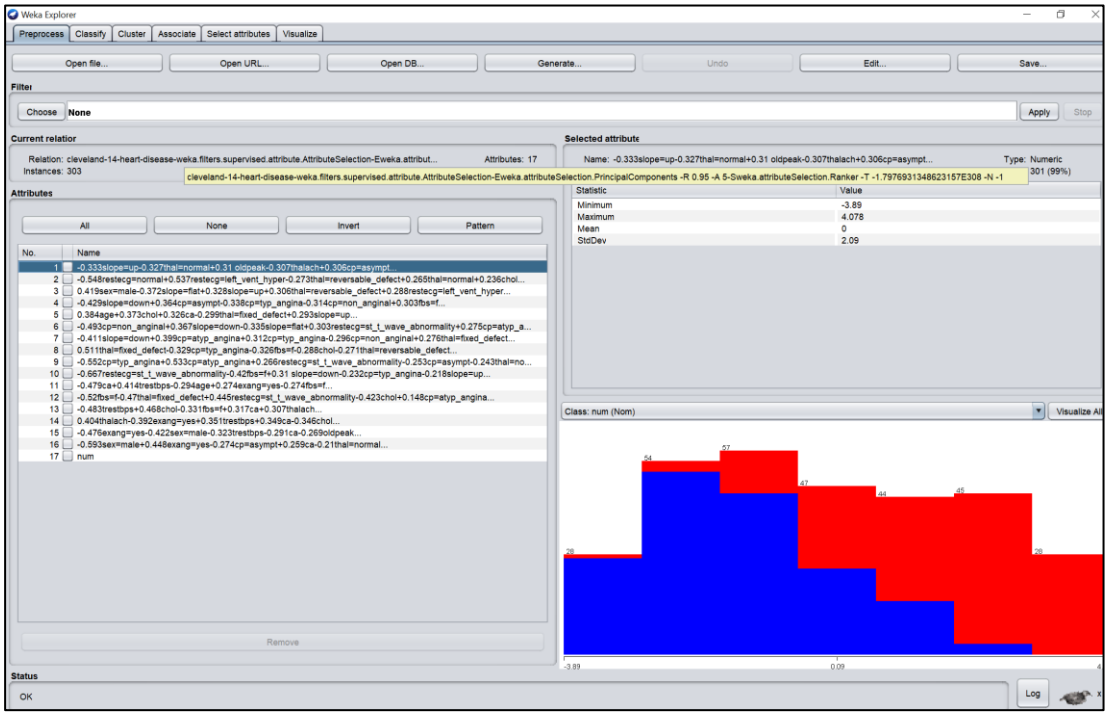

Figure 19: After applying PCA feature selection

Table 10: Classifiers performance without feature selection

Evaluation criteria	KNN	SMO	Logistic regression	Naive Bayes	MP	REP tree	Random forest	KStar	Bagging
Time to build model (s)	0	0.11	0.29	0.01	1.46	0.05	0.34	0	**0**
Correctly classified instances	231	255	256	253	245	232	252	226	**246**
Incorrectly classified instances	72	48	47	50	58	71	51	77	**57**

Table 11: Classifiers performance after applying feature selection method PCA

Evaluation criteria	KNN	SMO	Logistic regression	Naive Bayes	MP	REP tree	Random forest	KStar	Bagging
Time to build model (s)	0	0.02	0.08	0	0.87	0.01	0.29	0	**0.03**
Correctly classified instances	219	257	255	253	251	231	248	216	**252**
Incorrectly classified instances	84	46	48	50	52	72	55	87	**51**

Table 10 shows that logistic regression classifier has the best results. However, after applying PCA feature selection method as shown in *Table 11*, SVM has the best results. Therefore, the method that we will use in heart disease prediction system will support vector machine combined with PCA feature selection to deliver the best results.

5.2.2.2 Classification:

In this stage, we experiment under Weka environment all the classification methods discussed in this project for the purpose of comparative analysis. We created 10 classification models in Weka Explorer, a model for each classifier like support vector machine (SMO), K-Nearest Neighbors method (KNN), logistic regression, random forest, neural network (MP), bagging, K-star, Naïve Bayes and RIPPER with cross-validation 10 folds. We compared the accuracy of the classifiers to choose the best classifier for our dataset to use it to build the final prediction model. *Table 12* show the

accuracy for each classifier measured based on TP Rate, FP Rate, Precision, Recall, F-Measure and class. The accuracy results are shown before and after applying the PCA feature selection. As shown in *table 12* Support vector machine (SMO) has the best result which means building the prediction model using our dataset and SMO classifier will provide the best prediction results.

Table 12: Classifiers without feature selection

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
KNN	0.782	0.261	0.782	0.782	0.782	Absence
	0.739	0.218	0.739	0.739	0.739	Presence
SMO	0.897	0.225	0.827	0.897	0.860	Absence
	0.775	0.103	0.863	0.775	0.817	Presence
Logistic regression	0.891	0.210	0.835	0.891	0.862	Absence
	0.790	0.109	0.858	0.790	0.823	Presence
Naive Bayes	0.867	0.203	0.836	0.867	0.851	Absence
	0.797	0.133	0.833	0.797	0.815	Presence
MP	0.824	0.210	0.824	0.824	0.824	Absence
	0.790	0.176	0.790	0.790	0.790	Presence
REP tree	0.824	0.304	0.764	0.824	0.793	Absence
	0.696	0.176	0.768	0.696	0.730	Presence
Random forest	0.855	0.196	0.839	0.855	0.847	Absence
	0.804	0.145	0.822	0.804	0.813	Presence
KStar	0.800	0.319	0.750	0.800	0.774	Absence
	0.681	0.200	0.740	0.681	0.709	Presence
Bagging	0.885	0.275	0.793	0.885	0.837	Absence
	0.725	0.115	0.840	0.725	0.778	Presence
KNN	0.752	0.312	0.743	0.752	0.747	Absence
	0.688	0.248	0.699	0.688	0.693	Presence

Table 13: Classifiers after feature selection

	TP Rate	FP Rate	Precision	Recall	F-Measure	Class
SMO	0.897	0.210	0.836	0.897	0.865	Absence
	0.790	0.103	0.865	0.790	0.826	Presence
Logistic regression	0.897	0.255	0.827	0.897	0.860	Absence
	0.775	0.103	0.863	0.775	0.817	Presence
Naive Bayes	0.873	0.210	0.832	0.873	0.852	Absence
	0.790	0.127	0.838	0.790	0.813	Presence
MP	0.867	0.217	0.827	0.867	0.846	Absence
	0.783	0.133	0.831	0.783	0.806	Presence
REP tree	0.818	0.304	0.763	0.818	0.789	Absence
	0.696	0.182	0.762	0.696	0.727	Presence
Random forest	0.867	0.239	0.813	0.867	0.839	Absence
	0.761	0.133	0.827	0.761	0.792	Presence
KStar	0.739	0.319	0.735	0.739	0.737	Absence
	0.681	0.261	0.686	0.681	0.684	Presence
Bagging	0.879	0.225	0.824	0.879	0.850	Absence
	0.775	0.121	0.843	0.775	0.808	Presence

5.2.2.3 Building the prediction model:

After experimenting the classifiers in Weka and applying feature selection method PCA, SVM classifier showed the best results. Thus, we will implement the prediction model using SVM classifier in python programming language using PyCharm IDE and Anaconda.

- First, we imported all needed libraries as shown in figure 20.

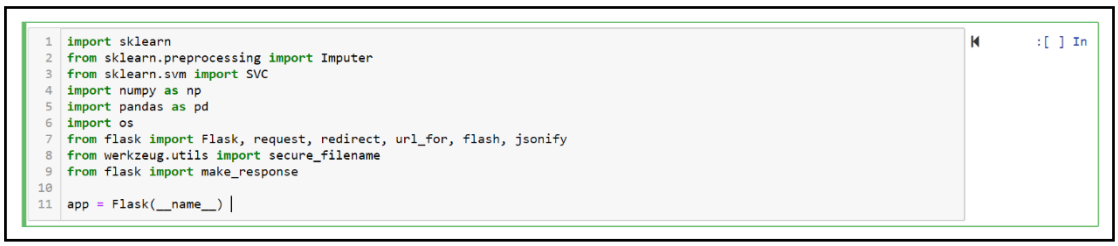

Figure 20: SVM model

To implement SVM classifier we need to import Sklearn as shown in figure 20, line 1. Sklearn is a shortcut for Scikit-learn. Scikit-learn is a library in python which provides many supervised and unsupervised algorithms and focus on modeling data. We imported the SVM classifier from this library as shown in figure 20line 3.[10]
The functionalities that Sklearn library provides include:

- Classification
- Regression
- Clustering
- Preprocessing

NumPy and Pandas libraries:

NumPy is shown in *figure 20*, line 4. NumPy is the core library for scientific computing in python. It provides high-performance multidimensional array objects and tools to work with the array.

Pandas as shown in *figure 20*, line 5 is an open-source library that provides high-performance, easy-to-use data structures and data analysis tools for the Python programming language. It allows us to import and analyze our data.

Flask framework:

Flask is web application framework in python. We use it to integrate our model with our website [43].

After importing all needed libraries, we imported our dataset as shown in *figure 21*, line 21 which contains nominal or string data. W performed categorization to transform the string data to numeric data as shown in *figure 21*, from line 12 to line 19. Following, we defined an array containing all data columns in the dataset except the last data column which is the diagnose as shown in *figure 21*, line 24. Moreover, we defined a for loop which will check every data column and transform the string values to the equivalent defined numeric values as shown in *figure 21*, from the line 28 to the line 33.
Finally, we defined the method for web applications @app as shown in figure 21, line 39. This method will be given inputs entered by the user in the prediction form on our website. The user will input string and numeric values. Subsequently, the method @app

will transform the string data to numeric as shown in *figure 21*, line 48 then preform the SVM classifier and predict the answer based on the dataset as shown in line 55 to line 59. The answer will be in array/matrix format. The method will take the first value in the array which is the predicted answer and send it back to the website. The website code will take the answer and visualize it to the user as negative or positive.

```
app = Flask(__name__)
gender_dict = {'male' : 0, 'female': 1}
cp_dict = {'asympt' : 0, 'non_anginal': 1, 'typ_angina':2, 'atyp_angina':3}
fbs_dict = {'f' : 0, 't' : 1}
restecg_dict = {'normal':0, 'left_vent_hyper':1, 'st_t_wave_abnormality':2}
exang_dict = {'yes' : 0, 'no' : 1}
slope_dict = {'up': 0, 'flat':1, 'down':2}
thal_dict = {'normal':0, 'reversable_defect':1, 'fixed_defect':2}
tr = {'1' : gender_dict, '2' : cp_dict, '5' : fbs_dict, '6' : restecg_dict, '8' : exang_dict, '10' : slope_dict, '12' :

x = pd.read_csv("x.csv")
a = np.array(x)
y  = a[:,13]
x = np.column_stack((x.age,x.sex,x.cp,x.trestbps,x.chol,x.fbs,x.restecg,x.thalach,x.exang,x.oldpeak,x.slope,x.ca,x.thal))

clf = SVC(kernel='linear')

for row in x:
    for col_idx in range(len(row)):
        column = row[col_idx]
        if str(col_idx) in tr:
            d = tr[str(col_idx)]
            row[col_idx] = d[row[col_idx]]


clf.fit(x, y)


@app.route('/predict', methods=['GET', 'POST'])
def predict():
    if request.method == 'GET':

        data = [[request.args.get('par1'),request.args.get('par2'),request.args.get('par3'),request.args.get('par4'),
                 request.args.get('par5'),request.args.get('par6'),request.args.get('par7'),request.args.get('par8'),
                 request.args.get('par9'),request.args.get('par10'),request.args.get('par11'),request.args.get('par12'),
                 request.args.get('par13')]]

        for row in data:
            for col_idx in range(len(row)):
                column = row[col_idx]
                if str(col_idx) in tr:
                    d = tr[str(col_idx)]
                    row[col_idx] = d[row[col_idx]]

        answer = clf.predict(data)

        return make_response(jsonify({
            'answer' : answer[0]
        }))

if __name__ == '__main__':
    app.secret_key = '12345678'
    app.make_null_session()
    app.run(debug=True)
```

Figure 21:SVM model

Figure 22 shows the code that integrates the user inputs in the make prediction form on our website with the model. The code sends the user inputs to the prediction model then the @app method in the model will execute the inputs and predict the answer. The answer will be translated to negative or positive as shown in *figure22* line 59 to 62. If the prediction answer was >50_1, that means the risk of developing a heart disease is high (positive). Otherwise, the risk of devolving a heart disease is very low (negative).

```php
$answer = '';

if (isset($_POST['make_prediction'])) {

    $par1 = $_POST['age'];
    $par2 = $_POST['sex'];
    $par3 = $_POST['cp'];
    $par4 = $_POST['trestbps'];
    $par5 = $_POST['chol'];
    $par6 = $_POST['fbs'];
    $par7 = $_POST['restecg'];
    $par8 = $_POST['thalach'];
    $par9 = $_POST['exang'];
    $par10 = $_POST['oldpeak'];
    $par11 = $_POST['slope'];
    $par12 = $_POST['ca'];
    $par13 = $_POST['thal'];

    $request_url = "http://localhost:5000/predict?par1=$par1&par2=$par2&par3=$par3&par4=$par4
    &par5=$par5&par6=$par6&par7=$par7&par8=$par8&par9=$par9&par10=$par10&par11=$par11&par12=$par12&par13=$par13";
    //print $request_url;
    $response = json_decode(file_get_contents($request_url));
    $answer = $response->answer;

    if ($answer == '>50_1') {
        $answer = '+';
    } else {
        $answer = '-';
    }

    $sql = "insert into prediction values(null, '{$_SESSION['user_id']}', '$par1', '$par2', '$par3', '$par4', '$par5',
    '$par6', '$par7', '$par8', '$par9', '$par10', '$par11', '$par12', '$par13', '$answer')";
    mysqli_query($conn, $sql);

}
```

Figure 22: Integration with the model code

5.2.2.4 Interface implementation:

Our objective is to deliver a website that predicts the presence of a heart disease. The user will register/login to the website then enter the values of the features in the prediction form and enter predict. The model will run and make a prediction, the result will be displayed to the user.

I/O Screens

In this part we will provide screenshots for the main I/O screens

Home page

The user will enter the home page and login to the EHDP (effective heart disease prediction) website. If the user is a new user, he/ she will move to the register page by clicking the 'Register here' link.

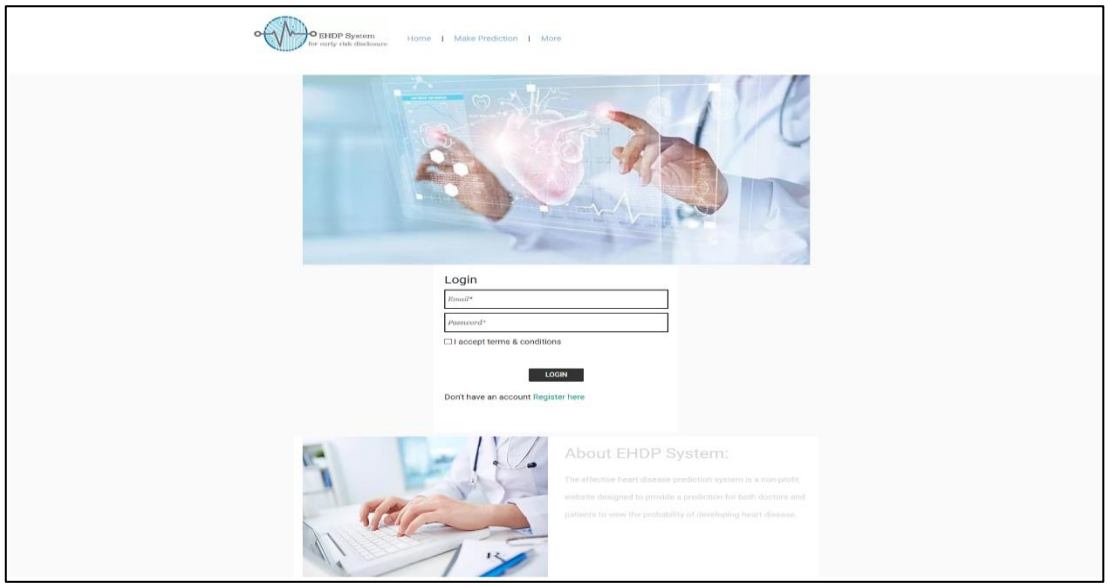

Figure 23:Screenshot of Home page

Register page:

The user can register to the website by filling the registration form and login by the Email and Password whenever using the website.

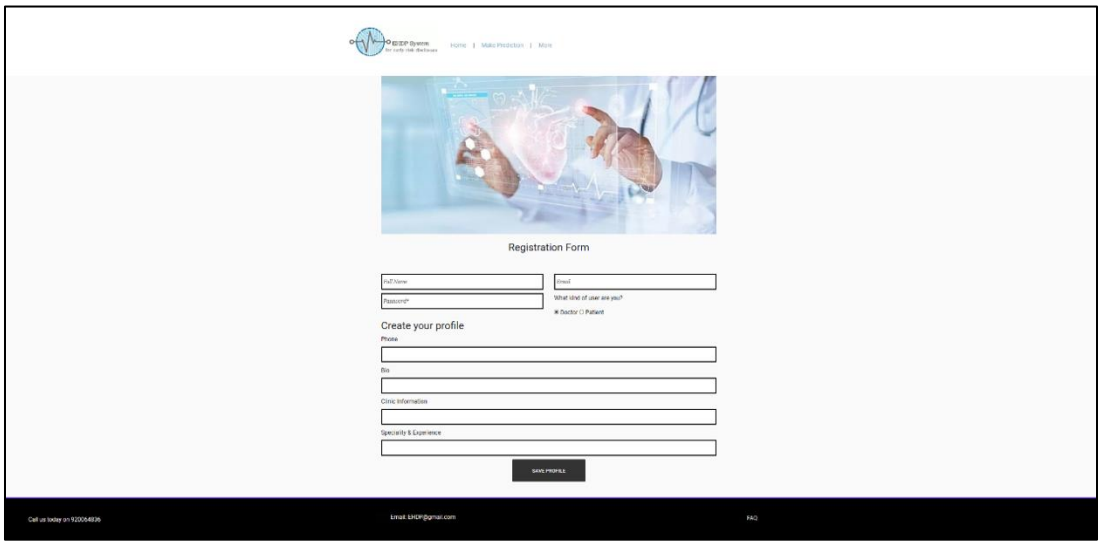

Figure 24:Screenshot of register page

Make prediction page:

After logging/ registering to the website the user can visit the make prediction page to predict the occurrence of a heart disease. The user needs to fill the prediction form and click predict.

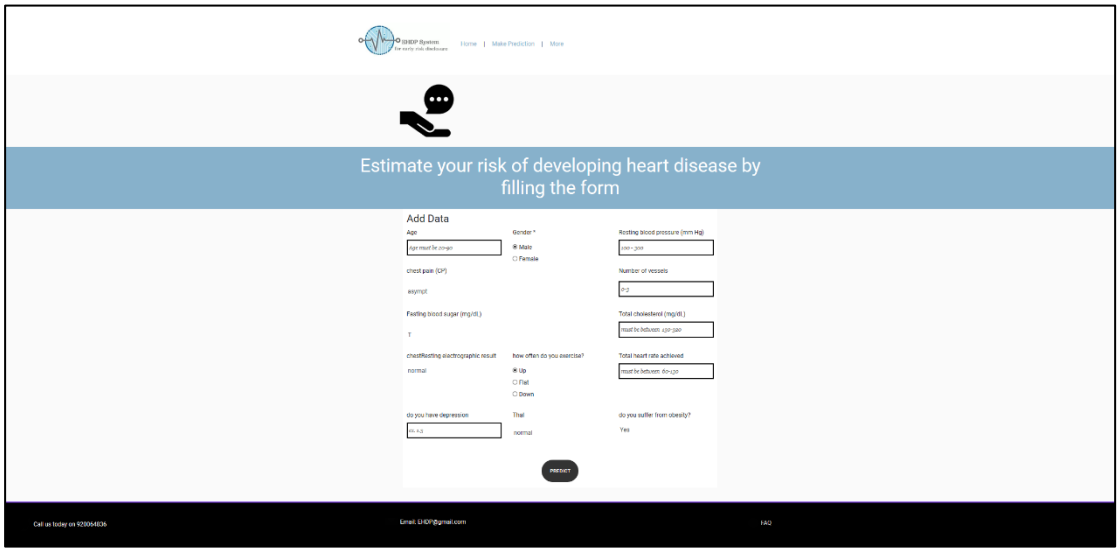

Figure 25: Screenshot of the make prediction page

Afterward, the result of the prediction will be displayed to the user in the same page in a box as shown in figure 30.

Figure 26: Result in the make prediction page

Tips page:

This page contains tips directed to patients to help them to manage their condition.

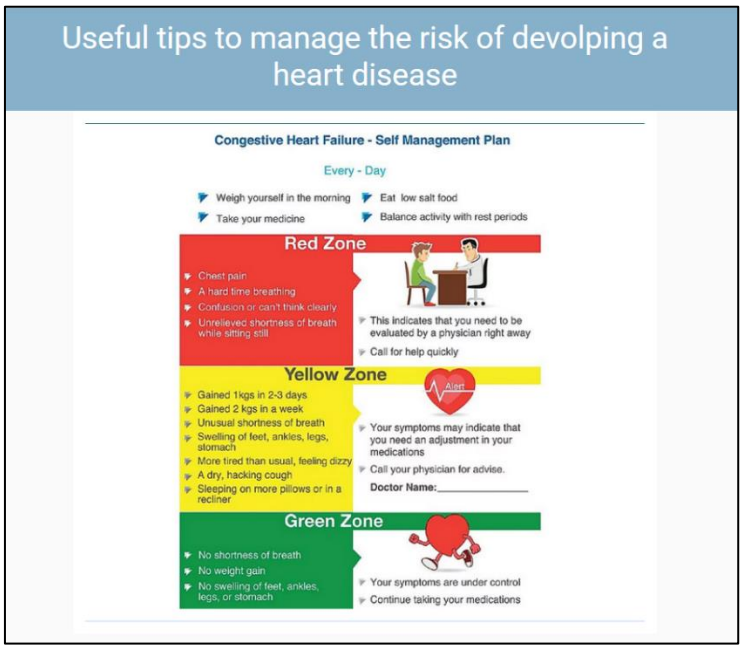

FIGURE 27: SCREENSHOT OF HELPFUL TIPS PAGE

Facts page:

This page provides statistics and facts about various heart diseases to increase awareness.

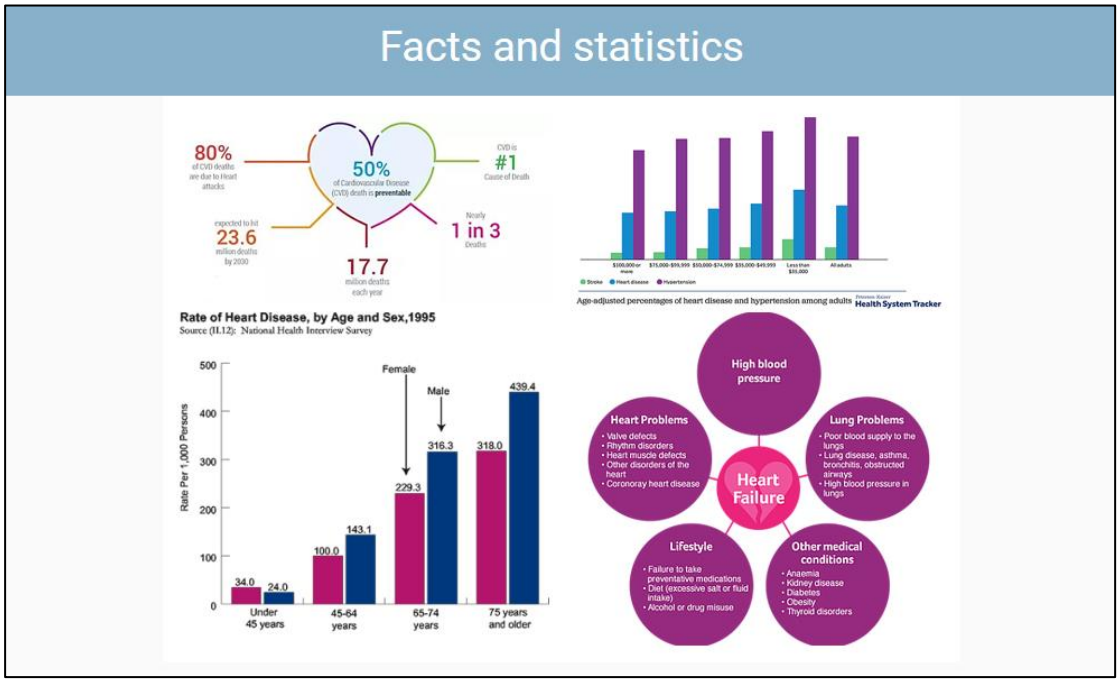

FIGURE 28: SCREENSHOT OF FACTS PAGE

View Nearest Hospital Page:

This page will display a map to show the users the nearest hospitals based on their location.

FIGURE 29: SCREENSHOT OF VIEW NEAREST HOSPITALS PAGE

FAQ page:

This page answers the users common questions related to the website and the EHDP system.

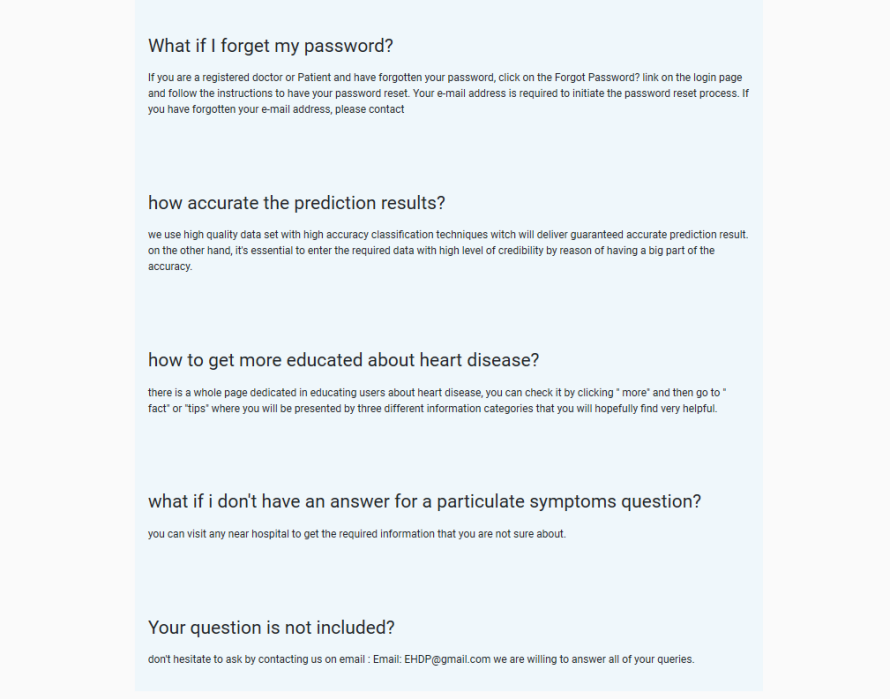

FIGURE 30: SCREENSHOT OF FAQ PAGE.

Chapter 6

6. Testing

Testing is the process of evaluating the system and its components. It is considered a primary step for any technical project. Therefore, we have followed some strategies to test the system in order to make sure that there is no system flaws, errors or gaps.

Test Objectives:

- Identify any errors may face the user in the future.
- Ensure the system is running smoothly and is bugs free.
- Discover the difficulties the user might encounter while using the website.
- Measure the quality of the system and find if the requirement matches the implementation.
- Test troubleshooting and maintenance issues that may arise after deployment.

Test Deliverables:

1. Test plan.
2. Test cases.
3. Tests execution.
4. Tests evaluation.
5. Documented tests results.

6.1 Testing plan

Our testing plan had to be comprehensive in a way that it evaluates all the specified requirements and ensures that the system met those requirements, and whether the system makes predictions properly without any errors or not.

The plan is to apply 4 testing levels including white and black box testing:

White Box Testing:

White Box testing is a technique that requires internal knowledge of the system. White box testing involves evaluating the system code and the internal structure of it. It aims to verify the flow of inputs and outputs through the system, including improving design, usability and security.

Black Box Testing:

Black-box testing is an important technique that does not require internal knowledge of the system. It is the method of system testing that examines and monitors the functionality of the system based on the pre-decided specifications and functional requirements.

Unit Testing:

Unit testing is a technique where individual modules are tested by the developer to determine if there are any issues with the standalone modules. The main goal is to isolate each unit then identify, analyze and fix the errors.

Integration Testing:

It is a process where the individual units are combined and tested as a group. The purpose is to expose flaws in the interaction between the integrated units, and to verify the functional, performance, and reliability between the integrated modules.

System Testing:

System Testing is a black box testing technique performed to evaluate the complete integrated system. The purpose of this process is to evaluate the system's compliance with the specified requirements. It includes both functional and Non-Functional testing.

Acceptance Testing:

Acceptance testing is a process where a system is tested for acceptability. The purpose of this test is to evaluate the system's compliance with the business requirements then assess and verify whether it is acceptable for delivery to end-users.

6.1.1 Schedule of Test Activities:

A test schedule includes testing levels, the testing activity or tasks and the time that the testing activity needs in order to be completed. Test activities scheduling is a crucial step in any testing plan. It specifies what should be tested and the duration of that test activity which helps the testing team with time management and to be focused on the specified tasks.

Table 14: Database Test cases

Testing Level	Test Activity	Duration
Unit Testing	Test and validate website registering data before submitting	1 Week
	Test and validate website logging data	
	Test and validate the retrieval of user's profile information from database.	
	Test and validate update user's profile information in database.	
	Test and validate the SVM classifier Prediction Model	
Integration Testing	Test and validate the prediction model after integrating with the website	3 Days
System Testing **(Functional)**	Testing and validate all users' features	2 Days
	Testing and validate all the website functions	
System Testing **(Non-Functional)**	Test the website interface to achieve an acceptable level of ease of use and consistency	1 Day
Acceptance Testing	Test the interface appearance and database performance in addition to the prediction model in order to achieve acceptable results	1 Day

6.2 Test Cases

In this section, we will present the test cases that have been conducted to ensure that the system works in a proper way and the proposed solution in case of any failure.

6.2.1 Database Test Cases

Database is an important part of the system and we have to test every entry in the database and make sure that is correct.

Table 15: Database Test cases

Test Case	Pass/Fail	Type of raised error	Attempted Solutions
Table Names	Pass	No Errors	Table name were set correctly
Field Names	Pass	No Errors	Field name were set correctly
Data Types	Pass	No Errors	Every data type is suitable for its field name
Size of Each Attribute	Pass	No Errors	Every size is suitable for its field name
Primary Keys	Pass	No Errors	Primary keys were set correctly
Foreign Keys	Pass	No Errors	Foreign keys between tables were set correctly

6.2.2 Functional Test Cases

This part concentrates on verifying that the system provides the expected functionalities to do so, we tested the following:

User Test Cases

User test cases help guide the tester through a sequence of steps to validate whether the system is error-free, and it is working as required by the end-user.

Table 16: User Test cases

Test Case	Pass/Fail	Type of raised error	Attempted Solutions
User can open the website.	Pass	No Errors	The user can open the website successfully
User can create an account and log in.	Pass	No Errors	The user can register and log in successfully
User can insert data into the form.	Pass	No Errors	The user can insert data successfully
User can view the prediction result.	Pass	No Errors	Prediction result is viewable to the user
User can view all pages.	Pass	No Errors	The user can view all pages successfully
User can log out of the website	Pass	No Errors	logging out is successful

6.2.3 Non-functional Test Cases

This part is to ensure that the project's non-function requirements are properly satisfied by the system. It concentrates on the way the system operates, rather than specific behaviors of that system. We tested the following:

User Interface Test Cases

The purpose of testing the graphical user interface (GUI) is to ensure that the user interface functionality works as per the specification.

Table 17: User Interface Test cases

Test Case	Pass/Fail	Type of raised error	Attempted Solutions
All the website required elements are available	Pass	No Errors	Success.
System is high performance and can handle throughput and response time.	Pass	No Errors	Success.
System is easy to lean and operate on	Pass	No Errors	Success
Unauthorized users are not able to log into the web interface.	Pass	No Errors	Success
Website pages are spelling and grammar correct	Pass	No Errors	Success

6.2.4 Acceptance testing

The final level of testing is acceptance testing, which evaluates the system's compliance with the real-life use and assesses whether it is acceptable for delivery. User Acceptance Testing (UAT) is the process of testing a software by the user to determine whether it can be accepted or not. This test will be carried out by the end-users who are familiar with the system requirements. The test was carried out by 4 individuals from PNU 3 of them are students and one doctor.

Table 18: Acceptance test

User Evaluation												
Users	Accessibility			Navigation			Response Time			Appearance		
	Good	Medium	Low	Good	Medium	Low	Good	Medium	Low	Good	Medium	Low
User 1	✓			✓				✓		✓		
User 2	✓			✓			✓				✓	
User 3	✓			✓			✓				✓	
User 4	✓				✓		✓			✓		

6.3 Test results.

After conducting the previous testing levels: unit testing, integration testing and system testing in addition to acceptance testing, we came to the conclusion that both the website and prediction model are error-free, can perform all function and achieved all the specified requirements.

Chapter 7

7. Conclusion:

The technology industry keeps growing and evolving every day and data mining applications are becoming more easier to use, broader, and more helpful. In our project, we developed a system that can detect heart diseases with maximum accuracy by mining and analyzing heart disease dataset. We performed the classification experiment using Weka and PyCharm software to execute many data mining techniques passing through all the development processes from considering the problem to gathering requirements ending with the needed system. As a result of the experiment, we created a prediction model using support vector machine data mining technique to predict the occurrence of a heart disease. In odreder to build the model, the system used 14 medical features for prediction such as age, gender, fasting blood sugar, blood pressure, cholesterol, and obesity. Finally, we developed a website which can be used by patients and doctors to diagnose heart disease.

7.1 Evaluation:

Passing through all of the previous processes, from analyzing the data, applying the classification algorithms, enhancing them with feature selection method, building the model, and integrate the model with the website, we ended up with a system that is capable to predict if a heart disease would occur or not. Also, the development of our project helped us to improve our programming skills, research, and teamwork skills.

7.2 Future work:

Heart diseases is significant and sensitive topic and its occurrence worries both doctors and patients. For this matter, we provided a website that predict the occurrence of a heart disease which can help to get an insight into the future of health examinations. Overall, this website was built on data mining experiments.

For future work, there are many ideas that could be applied and tested.

A. We would like to implement the system using data from KSA.
B. The prediction could be enhanced to predict the risk level of the disease (requires more accurate data).

C. The prediction system could be enhanced to give scores of different heart disease and the percentage of occurrence of each disease.

D. In the future, we would like to predict the time of occurrence, which will give more helpful insights.

8. References

[1] M. o. Health, "World Heart Day," 29 sptember 2018. [Online]. Available: https://www.moh.gov.sa/en/HealthAwareness/HealthDay/2018/Pages/HealthDay-2018-09-29.aspx. [Accessed 13 october 2019].

[2] u. i. a. t. force, "The Investment Case for Noncommunicable Disease Prevention and Control in the Kingdom of Saudi Arabia," 1 june 2017. [Online]. Available: https://www.undp.org/content/dam/saudi_arabia/docs/Publications/180326%20MOH%20KSA%20NCDs%202017.pdf. [Accessed 15 october 2019].

[3] M. F. M. Ghani, "Intelligent heart disease prediction system using data mining techniques," march 2008. [Online]. Available: https://www.researchgate.net/publication/4329399_Intelligent_heart_disease_prediction_system_using_data_mining_techniques. [Accessed 15 october 2019].

[4] M. &. F. S. Waller, "Data Science, Predictive Analytics, and Big Data: A Revolution that Will Transform Supply Chain Design and Management," 2 november 2013. [Online]. Available: https://www.academia.edu/22119017/Data_Science_Predictive_Analytics_and_Big_Data_A_Revolution_That_Will_Transform_Supply_Chain_Design_and_Management. [Accessed 15 october 2019].

[5] towardsdatascience, "Heart Disease Prediction," [Online]. Available: https://towardsdatascience.com/heart-disease-prediction-73468d630cfc. [Accessed 20 october 2019].

[6] P. Singh, S. Singh and G. S. Pandi-Jain, "Effective heart disease prediction system using data mining techniques," *International Journal of Nanomedicine,* p. 121–124, 2019.

[7] ministry of health, "Cardiovascular Diseases Cause 42% of Non-Communicable Diseases Deaths in the Kingdom," 30 october 2013. [Online]. Available: https://www.moh.gov.sa/en/Ministry/MediaCenter/News/Pages/News-2013-10-30-002.aspx. [Accessed 25 october 2019].

[8] Mayo Clinic Staff, "Heart disease," 22 march 2018. [Online]. Available: Heart disease describes a range of conditions that affect of heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease; heart rhythm problems (arrhythmias); heart attack; congestive heart failure, and . [Accessed 25 october 2019].

[9] A. A. S. W. L. M. E. A. A. A.-M. Hassan Kalaf, "Cardiovascular disease risk profile among young Saudi women of Al-Qassim, Saudi Arabia: A cross-sectional study," *international journal of health sciences,* january 2016.

[10] D. Alexander, "Data Mining," The University of Texas at Austin, Texas, 2019.

[11] C. Z. Charu C. Aggarwal, in *Mining Text Data*, new york, Springer Science & Business Media, 2012, p. 524.

[12] D. Alexander, "Data Mining," The University of Texas at Austin, Texas, 2019.

[13] microstrategy, "Data Mining Explained," [Online]. Available: https://www.microstrategy.com/us/resources/introductory-guides/data-mining-explained. [Accessed 16 october 2019].

[14] S. Alam, in *Biologically-Inspired Techniques for Knowledge Discovery and Data Mining*, IGI Global, 2014, p. 375.

[15] M. K. a. J. P. Jiawei Han, "Data Mining," in *Concepts and Techniques*, USA, Morgan Kaufmann, 2012, p. 744.

[16] R. D. Peng, "Leanpub," in *R Programming for Data Science*, Lean Publishing, 2015, p. 184.

[17] I. W. E. F. M. H. C. Pal, "Practical Machine Learning Tools and Techniques," in *Data Mining*, San Francisco, Morgan Kaufmann, 2016, p. 654.

[18] H. B. F. D. a. S. A. Belcy, "HEART DISEASE PREDICTION USING DATA MINING TECHNIQUES," Department of Computer Science and Engineering, Manonmaniam Sundaranar University, india, 2018.

[19] P. S. ,. G. S. P. Sanjay Singh, "Effective heart disease prediction system using data mining techniques," International Journal of Nanomedicine, india, 2018.

[20] S. K. H. S. G. K. S. A. a. Selvakumar, "Analysis of the Risk Factors of Heart Disease Using Step-Wise Regression with Statistical Evaluation," Springer, Cham, Switzerland , 2019.

[21] S. S. A. P. Chaitrali S. Dangare, "Improved Study of Heart Disease Prediction System using Data Mining Classification Techniques," International Journal of Computer Applications, india, 2012.

[22] N. AL-MILLI, "BACKPROPOGATION NEURAL NETWORK FOR PREDICTION OF HEART DISEASE," Journal of Theoretical and Applied Information Technology, Jordan, 2013.

[23] E. E. T. Santhanam, "Heart Disease Prediction Using Hybrid Genetic Fuzzy Model," Indian Journal of Science and Technology, india, 2015.

[24] J. Thomas and R. T. Princy, "Human heart disease prediction system using data mining techniques," IEEE, india, 2016.

[25] A. S. A. Dekamin, "A Data Mining Approach for Coronary Artery Disease Prediction in Iran," Journal of Advanced Medical Sciences and Applied Technologies, iran, 2017.

[26] S. S. T R Stella Mary, "Predicting heart ailment in patients with varying number of features using data mining techniques," International Journal of Electrical and Computer Engineering , india, 2019.

[27] M. F. M. Ghani, "Web-Based Heart Disease Decision Support System using Data Mining Classification Modeling Techniques," The Ninth International Conference on Information Integration and Web-based Applications Services, ndonesia, 2007.

[28] J. Soni, U. Ansari and D. Sharma, "Intelligent and Effective Heart Disease Prediction System using Weighted Associative Classifiers," International Journal on Computer Science and Engineering, india, 2011.

[29] M. R. R. M. R. I. M. H. I. Shadman Nashif, "Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System," World Journal of Engineering and Technology,, Bangladesh, 2018.

[30] D. P. a. J. M. P. Leonardo Yunda PhD, "A Web-based Fuzzy Inference System Based Tool for Cardiovascular Disease Risk Assessment," Universidad Colegio Mayor de Cundinamarca, Colombia, 2015.

[31] american college of cardiology, "Risk Estimator Plus," american college of cardiology, 1949. [Online]. Available: https://tools.acc.org/ASCVD-Risk-Estimator-Plus/#!/calculate/estimate/. [Accessed 21 november 2019].

[32] mdcalc, MDCalc, 2005. [Online]. Available: https://www.mdcalc.com/grace-acs-risk-mortality-calculator. [Accessed 12 november 2019].

[33] MDCalc , 2005. [Online]. Available: https://www.mdcalc.com/. [Accessed 12 november 2019].

[34] C. Althoff, "PyCharm," 2020. [Online]. Available: https://www.jetbrains.com/pycharm/.

[35] M. Sandmaier, "National Institutes of Health," 1 november 2005. [Online]. Available: https://www.nhlbi.nih.gov/files/docs/public/heart/living_well.pdf. [Accessed 13 october 2019].

[36] J. P. M. K. Jiawei Han, "Data Mining," in *Data Mining: Concepts and Techniques*, USA, morgan kaufmann, 2012, p. 774.

[37] 2016. [Online]. Available: https://ieeexplore-ieee-org.sdl.idm.oclc.org/stamp/stamp.jsp?tp=&arnumber=7530265. [Accessed 18 october 2019].

[38] K. R. G. &. G. A. A. J. S. E. Srinivas, "Rough-Fuzzy Classifier: A System to Predict the Heart Disease by Blending Two Different Set Theories," *Arabian Journal for Science and Engineering*, Vols. 2191-4281, 9 February 2014.

[39] general authority for statistics, "Household Health Survey," 2018. [Online]. Available: https://www.stats.gov.sa/sites/default/files/household_health_survey_2018.pdf. [Accessed 16 october 2019].

[40] RStudio, "RStudio," 2019. [Online]. Available: https://rstudio.com/about/. [Accessed 16 october 2019].

Appendices:

A. Interview questions:

السلام عليكم ورحمة الله وبركاته

نحن طالبات من جامعة الأميرة نورة بنت عبد الرحمن، كلية علوم الحاسب والمعلومات تخصص نظم معلومات، نعمل حالياً على مشروع التخرج والذي يهدف الى تقديم وتطوير نظام فعال للتنبؤ بأمراض القلب، احتمالية الإصابة بالمرض ومستوى خطورة الإصابة.
يهدف المشروع الى مساعدة الأطباء في تشخيص الأمراض بدقة عالية وفعالة، يساعد ايضاً المرضى في تتبع حالتهم الصحية واحتمالية اصابتهم بالمرض.

نأمل منكم مساعدتنا في الإجابة على بعض الأسئلة لمساعدتنا في جمع البيانات اللازمة:

1. ماهي اعراض الإصابة بأمراض القلب؟
2. ماهي العوامل المؤثرة التي تعزز الإصابة بأمراض القلب؟
3. بنظرك، ماهي الوسائل التي تسهل التواصل بين المريض وطبيب القلب؟
4. هل يوجد انظمه مستخدمه من قبل الأطباء للمساعدة بالتنبؤ في أمراض القلب وتحديد خطورتها؟
5. ماهي الخدمات التي ترجو توفرها في النظام؟
6. هل يوجد لديكم اهتمام في استخدام النظام المذكور أعلاه؟

B. Database code:

```sql
create table users
(
    id          int primary key auto_increment,
    full_name   varchar(50),
    email       varchar(30) unique,
    password    text,
    user_type   int default 0 comment '0 patient, 1 doctor',
    phone       varchar(20),
    bio         text null,
    clinic_info text null,
    experience  text null
);

create table prediction
(
    id       int primary key auto_increment,
    user_id  int,
    age      int,
    sex      varchar(20),
    cp       varchar(100),
    trestbps int,
    chol     int,
    fbs      varchar(100),
    restecg  varchar(100),
    thalach  int,
    exang    varchar(20),
    oldpeak  float,
    slope    varchar(20),
    ca       int,
    thal     varchar(100),
    result   varchar(100),

    constraint foreign key (user_id) references users (id)
        on delete cascade on update cascade
);
```

C. Facts page code:

```php
<?php
include_once 'header.php';
?>

<div class="container">
    <div class="row justify-content-center">
        <div class="col-md-10" style="padding: 20px;box-sizing: border-box;background-color: #eff7fb">

            <div>
                <br><br><br><br>
                <h3>What if I forget my password?</h3>
                <p style="margin-top: 20px;">
                    If you are a registered doctor or Patient and have forgotten your password, click on the Forgot Password?
                    link on the login page and follow the instructions to have your password reset. Your e-mail address is required to initiate the password reset process.
                    If you have forgotten your e-mail address, please contact
                </p>
            </div>

            <div>
                <br><br><br><br>
                <h3>how accurate the prediction results?</h3>
                <p style="margin-top: 20px;">
                    we use high quality data set with high accuracy classification techniques witch will deliver guaranteed accurate prediction result.
                    on the other hand, it's essential to enter the required data with high level of credibility by reason of having a big part of the accuracy.
                </p>
            </div>

            <div>
                <br><br><br><br>
                <h3>how to get more educated about heart disease?</h3>
                <p style="margin-top: 20px;">

                    there is a whole page dedicated in educating users about heart disease, you can check it by clicking " more" and then go to " fact" or "tips"
                    where you will be presented by three different information categories that you will hopefully find very helpful.
                </p>
            </div>



            <div>
                <br><br><br><br>
                <h3>what if i don't have an answer for a particulate symptoms question?</h3>
                <p style="margin-top: 20px;">
                    you can visit any near hospital to get the required information that you are not sure about.
                </p>
            </div>

            <div>
                <br><br><br><br>
                <h3>Your question is not included?</h3>
                <p style="margin-top: 20px;">
                    don't hesitate to ask by contacting us on email : Email: EHDP@gmail.com
                    we are willing to answer all of your queries.
                </p>
            </div>
        </div>
    </div>
</div>

<?php
include_once 'footer.php';
?>
```

D. FAQ page code

```
<?php
?>

<style>
    .footer {
        margin-top: 20px;
        padding: 40px;
        background-color: black;
        border-top: 2px solid #7133d7;
    }

    .footer a {
        color: #d7d7d7;
        text-decoration: none;
    }
    .footer a:hover {
        color: #d7d7d7;
    }
    .footer {
        color: white;
    }
</style>

<div class="footer">
    <div class="row">
        <div class="col-md-4">
            <img src="img/phone.jpg" style="width:32px;height:40px;object-fit:fill" alt="">
            Call us today on 920064836
        </div>
        <div class="col-md-4">
            <img src="img/email.jpg" alt="">
            Email: EHDP@gmail.com
        </div>
        <div class="col-md-4">
            <img src="img/faq.jpg" alt="">
            <a href="faq.php">FAQ</a>
        </div>
    </div>
</div>

<script src="https://code.jquery.com/jquery-3.2.1.slim.min.js"
        integrity="sha384-KJ3o2DKtIkvYIK3UENzmM7KCkRr/rE9/Qpg6aAZGJwFDMVNA/GpGFF93hXpG5KkN"
        crossorigin="anonymous"></script>
<script src="https://unpkg.com/popper.js@1.12.6/dist/umd/popper.js"
        integrity="sha384-fA23ZRQ3G/J53mElWqVJEGJzU0sTs+SvzG8fXVWP+kJQ1lwFAOkcUOysnlKJC33U"
        crossorigin="anonymous"></script>
<script src="https://unpkg.com/bootstrap-material-design@4.1.1/dist/js/bootstrap-material-design.js"
        integrity="sha384-CauSuKpEqAFajSpkdjv3z9t8E7RlpJ1UP0lKM/+NdtSarroVKu069AlsRPKkFBz9"
        crossorigin="anonymous"></script>
<script>$(document).ready(function () {
        $('body').bootstrapMaterialDesign();
    });</script>


<script>
    $(document).ready(function () {
        const dropDownMenu = $('.drop-down-menu');
        dropDownMenu.mouseenter(function () {
            const menu_id = $(this).data('show-menu');
            $(menu_id).css('display', 'block');
        });

        $('.custom-header').mouseleave(function () {
            // const menu_id = $(this).data('show-menu');
            $("#sub-menu-more").css('display', 'none');
        });
    });
</script>
</body>
</html>
```

E. Footer code


```
<?php
?>

<style>
    .footer {
        margin-top: 20px;
        padding: 40px;
        background-color: black;
        border-top: 2px solid #7133d7;
    }

    .footer a {
        color: #d7d7d7;
        text-decoration: none;
    }
    .footer a:hover {
        color: #d7d7d7;
    }
    .footer {
        color: white;
    }
</style>

<div class="footer">
    <div class="row">
        <div class="col-md-4">
            <img src="img/phone.jpg" style="width:32px;height:40px;object-fit:fill" alt="">
            Call us today on 920064836
        </div>
        <div class="col-md-4">
            <img src="img/email.jpg" alt="">
            Email: EHDP@gmail.com
        </div>
        <div class="col-md-4">
            <img src="img/faq.jpg" alt="">
            <a href="faq.php">FAQ</a>
        </div>
    </div>
</div>

<script src="https://code.jquery.com/jquery-3.2.1.slim.min.js"
        integrity="sha384-KJ3o2DKtIkvYIK3UENzmM7KCkRr/rE9/Qpg6aAZGJwFDMVNA/GpGFF93hXpG5KkN"
        crossorigin="anonymous"></script>
<script src="https://unpkg.com/popper.js@1.12.6/dist/umd/popper.js"
        integrity="sha384-fA23ZRQ3G/J53mElWqVJEGJzU0sTs+SvzG8fXVWP+kJQ1lwFAOkcUOysnlKJC33U"
        crossorigin="anonymous"></script>
<script src="https://unpkg.com/bootstrap-material-design@4.1.1/dist/js/bootstrap-material-design.js"
        integrity="sha384-CauSuKpEqAFajSpkdjv3z9t8E7RlpJ1UP0lKM/+NdtSarroVKu069AlsRPKkFBz9"
        crossorigin="anonymous"></script>
<script>$(document).ready(function () {
        $('body').bootstrapMaterialDesign();
    });</script>


<script>
    $(document).ready(function () {
        const dropDownMenu = $('.drop-down-menu');
        dropDownMenu.mouseenter(function () {
            const menu_id = $(this).data('show-menu');
            $(menu_id).css('display', 'block');
        });

        $('.custom-header').mouseleave(function () {
            // const menu_id = $(this).data('show-menu');
            $("#sub-menu-more").css('display', 'none');
        });
    });
</script>
</body>
</html>
```

F. Header code

```php
<?php
session_start();
$conn = mysqli_connect('localhost', 'root', '', 'db');
if (!$conn) die();

if (isset($_POST['register'])) {
    global $conn;
    $pass = md5($_POST['password']);
    $sql = "insert into users values (null, '{$_POST['full_name']}', '{$_POST['email']}', '$pass', {$_POST['user_type']}, '{$_POST['phone']}', '{$_POST['bio']}', '{$_POST['clinic_info']}', '{$_POST['experience']}')";
    if (mysqli_query($conn, $sql)) {
        $_SESSION['user_id'] = mysqli_insert_id($conn);
        header('Location: ./index.php');
    } else {
        print mysqli_error($conn);
    }
}

$err = '';
if (isset($_POST['login'])) {
    $email = $_POST['email'];
    $pass = md5($_POST['password']);

    global $conn;
    $sql = "select * from users where email = '$email' and password = '$pass' limit 1";
    $res = mysqli_query($conn, $sql);
    if (mysqli_num_rows($res) > 0) {
        $_SESSION['user_id'] = mysqli_fetch_assoc($res)['id'];
        header('Location: ./index.php');
    } else {
        $err = 'Email or/and password is wrong, try again.';
    }
}


$answer = '';

if (isset($_POST['make_prediction'])) {

    $par1 = $_POST['age'];
    $par2 = $_POST['sex'];
    $par3 = $_POST['cp'];
    $par4 = $_POST['trestbps'];
    $par5 = $_POST['chol'];
    $par6 = $_POST['fbs'];
    $par7 = $_POST['restecg'];
    $par8 = $_POST['thalach'];
    $par9 = $_POST['exang'];
    $par10 = $_POST['oldpeak'];
    $par11 = $_POST['slope'];
    $par12 = $_POST['ca'];
    $par13 = $_POST['thal'];

    $request_url = "http://localhost:5000/predict?par1=$par1&par2=$par2&par3=$par3&par4=$par4&par5=$par5&par6=$par6&par7=$par7&par8=$par8&par9=$par9&par10=$par10&par11=$par11&par12=$par12&par13=$par13";
    //print $request_url;
    $response = json_decode(file_get_contents($request_url));
    $answer = $response->answer;

    if ($answer == '>50_1') {
        $answer = '+';
    } else {
        $answer = '-';
    }

    $sql = "insert into prediction values(null, '{$_SESSION['user_id']}', '$par1', '$par2', '$par3', '$par4', '$par5','$par6', '$par7', '$par8', '$par9', '$par10', '$par11', '$par12', '$par13', '$answer')";
    mysqli_query($conn, $sql);
```

```
        padding: 5px;
        padding-left: 10px;
        padding-right: 10px;
        width: 100px;
        border: 0;
        outline: 0;
        cursor: pointer;
    }

</style>

</head>
<body>

<div class="custom-header">
    <div class="container">
        <div class="row">
            <div class="col-md-3">
                <img src="img/logo.jpg" alt="" style="width: 100%;vertical-align: center">
            </div>
            <div class="col-md-9">
                <div class="menu" style="border: 0px solid #333;margin-top: 45px;position:relative;">
                    <span><a href="index.php">Home</a></span>

                    <span style="margin-right: 15px;"> | </span>

                    <span><a href="prediction.php">Make Prediction</a></span>

                    <span style="margin-right: 15px;"> | </span>

                    <span><a class="drop-down-menu" href="#" data-show-menu="#sub-menu-more">More


                        </a></span>

                    <div class="sub-menu-more" id="sub-menu-more">
                        <div class="menu-item">
                            <a href="faq.php">FAQ</a>
                        </div>
                        <div class="menu-item">
                            <a href="register.php">Register</a>
                        </div>

                        <div class="menu-item">
                            <a href="hospitals.php">View nearest hospitals</a>
                        </div>

                        <div class="menu-item">
                            <a href="facts.php">Facts</a>
                        </div>

                        <div class="menu-item">
                            <a href="tips.php">Tips</a>
                        </div>
                    </div>

                </div>
            </div>
        </div>
    </div>
</div>
```

G. Home page code

H. Hospitals page code

```php
<?php
include_once "header.php";
?>

<style>
    .risk_div {
        padding: 20px;
        box-sizing: border-box;
        background-color: #87B1CB;
        color: white;
        text-align: center;

    }
</style>
<div class="risk_div">
    <div class="container">
        <h1 style="font-size: 50px !important;">
            Here is the nearest hospital based on your location
        </h1>
    </div>
</div>

<div class="container">
    <div class="row justify-content-center">
        <div class="col-md-10" style="padding: 10px;background-color: white;box-sizing: border-box">
            <iframe src="https://www.google.com/maps/embed?pb=!1m18!1m12!1m3!1d14950959.936356548!2d36.04154841571084!3d23.81373745186018!2m3!1f0!2f0!3f0!3m2!1i1024!2i768!4f13.1!3m
            width="100%" height="450" frameborder="0" style="border:0;" allowfullscreen=""></iframe>
        </div>
    </div>
</div>

<?php
include_once "footer.php";
?>
```

I. Prediction page code

J. Register code

```php
<?php
include_once 'header.php';

?>

<div class="container">

    <div class="row justify-content-center">
        <div class="col-md-10">
            <img src="img/home.jpg" alt="" style="width: 100%;">
        </div>
    </div>
    <div class="row justify-content-center" style="margin-top: 20px;">
        <div class="col-md-10">
            <h3 class="text-center">
                Registration Form
            </h3>
            <br><br>

            <form method="post" action="register.php">
                <div class="row">
                    <div class="col-md-6">
                        <input type="text" placeholder="Full Name" required name="full_name">
                    </div>
                    <div class="col-md-6">
                        <input type="email" placeholder="Email" required name="email">
                    </div>
                </div>

                <div class="row">
                    <div class="col-md-6">
                        <input type="password" name="password" required placeholder="Password* ">
                    </div>
                    <div class="col-md-6">
                        <p>What kind of user are you?</p>

                        <label for=""><input type="radio" name="user_type" value="1"> Doctor</label>
                        <label for=""><input type="radio" name="user_type" value="0" checked> Patient</label>
                    </div>
                </div>

                <div class="row">
                    <div class="col-md-12">
                        <h3>Create your profile </h3>
                    </div>

                    <div class="col-md-12">
                        <label for="">Phone</label>
                        <input type="text" name="phone" id="">
                    </div>

                    <div class="col-md-12">
                        <label for="">Bio</label>
                        <input type="text" name="bio" id="" style="width: 100%">
                    </div>

                    <div class="col-md-12">
                        <label for="">Clinic Information</label>
                        <input type="text" name="clinic_info" id="" style="width: 100%;">
                    </div>

                    <div class="col-md-12">
                        <label for="">Speciality & Experience</label>
                        <input type="text" name="experience" id="" style="width: 100%;">
                    </div>

                    <div class="text-center col-md-12">
                        <button class="btn" name="register" type="submit" style="width: 200px;padding: 20px;">
                            Save Profile
                        </button>
                    </div>
```

K. Style code

```html
<!doctype html>
<html lang="en">
<head>
    <meta charset="UTF-8">
    <meta name="viewport"
          content="width=device-width, user-scalable=no, initial-scale=1.0, maximum-scale=1.0, minimum-scale=1.0">
    <meta http-equiv="X-UA-Compatible" content="ie=edge">
    <title>Document</title>
    <link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Roboto:300,400,500,700|Material+Icons">
    <link rel="stylesheet"
          href="https://unpkg.com/bootstrap-material-design@4.1.1/dist/css/bootstrap-material-design.min.css"
          integrity="sha384-wXznGJNEXNG1NFsbm0ugrLFMQPWswR3lds2VeinahP8N0zJw9VWSopbjv2x7WCvX" crossorigin="anonymous">

    <style>
        .custom-header {
            padding: 50px;
            background-color: white;
        }

        .menu a {
            text-decoration: none;
            color: #87B1CB;
            transition: color 0.4s ease 0s;
            margin-right: 15px;
        }

        .sub-menu-more {

            width: 200px;
            padding: 10px;
            background-color: white;
            position: absolute;
            border: 1px solid #87B1CB;
            top: 25px;
            left: 230px;
            display: none;
            z-index: 10000;
        }

        .sub-menu-more .menu-item {
            margin: 3px;
            transition: linear .5s;
            padding: 5px;
        }

        .sub-menu-more .menu-item a {
            display: block;
        }

        .login-form {
            padding: 20px;
            background-color: white;
        }

        input[type='text'], input[type='number'], input[type='password'], input[type='email'] {
            border-radius: 0;
            font: italic normal normal 14px/1.4em georgia, palatino, 'book antiqua', 'palatino linotype', serif;
            -webkit-appearance: none;
            -moz-appearance: none;
            border-width: 2px;
            background-color: rgba(255, 255, 255, 1);
            box-sizing: border-box !important;
            color: #000000;
            border-style: solid;
```

L. Tips page code

```php
<?php
include_once 'header.php';
?>

<style>
    .risk_div {
        padding: 20px;
        box-sizing: border-box;
        background-color: #87B1CB;
        color: white;
        text-align: center;


    }
</style>
<div class="risk_div">
    <div class="container">
        <h1 style="font-size: 50px !important;">
            Useful tips to manage the risk of devolping a heart disease

        </h1>
    </div>
</div>

<div class="container">
    <div class="row justify-content-center">
        <div class="col-md-10" style="background-color: white;box-sizing: border-box;padding: 20px">
            <div class="row">
                <div class="col-md-12">
                    <img src="img/tips.PNG" alt="" style="width: 100%;">
                </div>

            </div>


        </div>
    </div>
</div>

<?php
include_once 'footer.php';
?>
```