INFORMATICS INSTITUTE OF TECHNOLOGY

UNIVERSITY OF WESTMINSTER

INFORMATICS INSTITUTE OF TECHNOLOGY

DEPARTMENT OF COMPUTING

(B.Eng.) in Software Engineering

Module Code: 5COSC021C

Software Development Group Project

Module Leader: Mr. Banuka Athuraliya

Heart Failure Prediction System

Team: **Codex**

1.	A. A. M. Abeysirwardana	:	2019450	:	w1761114
2.	O.V.K Gamage	:	20200188	:	w1809989
3.	T.S.C.Yapa	:	20200554	:	w1830243
4.	B.K.M.S. Rodrigo	:	2019360	:	w1761300
5.	S.F.M.Thariq	:	20200758	:	w1833602

Declaration

We declare that this report is our original work and no part of it has been published anywhere else in the past. Any work from other authors is duly referenced and acknowledged. The information and data given in the report are authentic to the best of our knowledge.

Abstract

Cardiovascular diseases or heart diseases are the main reasons for many deaths all around the world nowadays. So, there should be an accurate and reliable system to diagnose those heart diseases in real-time for proper treatments. The medical industry gathers a large amount of health care data of patients which unfortunately is not used in a proper manner to discover hidden data about diseases and make effective decisions to diagnose heart diseases and their risk levels. The main purpose of our project is to build a heart disease prediction system using machine learning and data science. For prediction purposes, patients should have a medical report which contains all the necessary information about diseases and ECG images. Heart failure has four stages which are stages A, B, C, and D. So, the patients can get a clear idea about their current stage and type of heart disease by using this system. After predicting the result of the disease, patients can see those results and the system recommends some cardiologists who can be contacted via the online platform.

Keywords

Heart disease, Machine learning techniques, Prediction, Accuracy

Acknowledgment

First of all, we want to thank God for giving us the knowledge, strength, and courage to face all the challenges and complete this thesis. As the same thanks to each & every one of our parents who guided us, worked hard to give us a good education in a stream we loved. Thanks for believing in us, giving us advice, having faith in us and being an encouragement for us.

We need to express our gratitude and thanks to our mentor, Mr. Banuka Athuraliya for his great support, guidance through our project. We are thankful to him for directing us in the right direction whenever we got stuck or had any doubt. Great thanks for each and every staff member of our SDGP module for giving us the knowledge & feedback every week for guiding us to achieve great results within our project.

Our friends who have helped us whenever we needed a helping hand, the academic and non-academic staff of IIT who provided support to us throughout the degree program, we are forever grateful for you all.

Last but not least, thanks to everyone else who helped us even with a word or encouraged us through difficult times.

Table of Contents

Declaration ..... **i**

Abstract ..... **ii**

Acknowledgment ..... **iii**

Abbreviations table ..... **x**

Chapter 1 - Introduction **1**

1.1 Chapter Overview 1

1.2 Problem Background 1

1.3 Problem Statement 4

1.7 Project Scope 6

1.7.1 In - Scope 6

1.7.2 Out - Scope 6

1.8 Rich Picture Diagram 7

1.9. Objectives 8

1.9.1 Research Objectives 8

1.9.2 Academic Objectives 8

1.9.3 Operational Objectives 8

1.10 Resource Requirements 9

Technology Stack 9

1.10.3 Data Requirements 10

1.11 Chapter Summary 11

Chapter 2 - Literature Review **12**

2.1 Chapter Introduction 12

2.2 Previous Proposed Solution Overview 12
2.3 Existing Work 13
2.3.1 Various Existing Works and Results 13
2.3.1 Reviews on Existing Research 18
2.4 Tools and techniques 19
2.4.1 Review of Tools & Techniques 19
2.4.2 Review of evaluation Techniques 21
2.5 Selected Approach 22
2.6 Chapter Summary 22
Chapter 3 - Methodology 23
3.1 Chapter Overview 23
3.2 Research Methodology 23
3.2.1 Types of Methodologies 23
3.2.2 Selecting the Research Methodology 24
3.3 Development Methodology 24
3.6.1 Types of Development Methodologies 24
3.4 Design methodology 26
3.4.1 Types of Design Methodologies 26
3.4.2 Selecting the Design Methodology 27
3.5 Evaluation methodology 27
3.8 Gantt chart diagram 31
3.9 Usage of Project Management and Collaboration Software In the project 32
3.10 Risks and Mitigation 33
3.11 Chapter Summary 34

Chapter 4 - System Requirements Specification (SRS) 35
4.1 Chapter Overview 35
4.2 Stakeholder Analysis 35
4.2.1 Onion Model 35
4.2.2 Stakeholder Descriptions 36
4.3 Selection of Requirement Elicitation Techniques/Methods 37
4.4 Discussion/ Analysis of Results 38
4.5 Use Case Diagram 41
4.6 Use Case Descriptions 42
4.7 Functional Requirements (with prioritization) 51
4.8 Non-Functional Requirements 53
4.9. Chapter Summary 53
Chapter 5: Social, Legal, Ethical and Professional Issues 54
5.1 Chapter Overview 54
5.2 SLEP Issues and Mitigation 54
5.2.1. Social Issues 54
5.2.2 Legal Issues 54
5.2.3 Ethical Issues 55
5.2.4 Professional Issues 55
5.4 Chapter Summary 55
Chapter 6: System Architecture & Design 56
6.1 Chapter Overview 56
6.3 System Design 58
6.3.1 Class Diagram 58

6.3.2 Sequence Diagram 59

6.3.3 UI Design 61

6.3.4 Process flow chart 66

6.4 Chapter Summary 67

References I

Appendix V

Work Breakdown Chart V

List of Figures

Chapter 1

Figure 1. 1: Different Stages of Heart Failure (Tripathi, Sullivan, Fan and Mehta, 2021) 2
Figure 1. 2: Rich Picture Diagram 7

Chapter 2

Figure 2. 1: Various Machine learning techniques with their accuracy levels 17
Figure 2. 2: example decision tree (Machine Learning Decision Tree Classification Algorithm - Javatpoint, no date) 20

Chapter 3

Figure 3. 1: Different Phases of Agile Project Management Methodology (Kanban Versus Scrum in Agile Software Development, 2022) 28
Figure 3. 2: Work Breakdown Structure 30
Figure 3. 3: Gantt chart diagram 31

Chapter 4

Figure 4. 1: Stakeholder Analysis Onion Model 35
Figure 4. 2: Use Case Diagram 41

Chapter 6

Figure 6. 1: System Architecture Design 57
Figure 6. 2: Class Diagram 58
Figure 6. 3: Sequence Diagram 60
Figure 6. 4: Loading Page
Figure 6. 5: Signup Page……. 61
Figure 6. 6: Login Page
Figure 6. 7: Home Page 62
Figure 6. 8: Side Bar
Figure 6. 9: User Profile 63
Figure 6. 10: Doctors List
Figure 6. 11: Prediction Data 64
Figure 6. 12: Upload ECG page 65
Figure 6. 13: Process Flow Chart 66

List of Tables

Chapter 1

Table 1. 1: Different Types of Heart Diseases (Marimuthu et al., 2018) 3
Table 1. 2: Feature Comparison Chart. 4
Table 1. 3: Hardware Requirements and Software Requirements 9
Table 1. 4: Technology Stack 9

Chapter 2

Table 2. 1 Comparative analysis between various existing works with used techniques and their accuracy levels (Almustafa, 2020) 17
Table 2. 2: Advantages and Disadvantages of used techniques (Almustafa, 2020) 22

Chapter 3

Table 3. 1: Reasons for not choosing the other models 26
Table 3. 2: Risks and mitigation 34

Chapter 4

Table 4. 1: Stakeholder Analysis Description 37
Table 4. 2: Discussion/Analysis Table 40
Table 4. 3: Use Case Table 1 42
Table 4. 4: Use Case Table 2 44
Table 4. 5: Use Case Table 3 45
Table 4. 6: Use Case Table 4 46
Table 4. 7: Use Case Table 5 47
Table 4. 8: Use Case Table 6 48
Table 4. 9: Use Case Table 7 49
Table 4. 10: Use Case Table 8 50
Table 4. 11: Functional requirements 52
Table 4. 12:Non-Functional requirements 53

Abbreviations table

Abbreviation	Description
SVM	Support Vector Machine Algorithm
FT	Functional Tree
GA	Genetic Algorithm
KNN	K-nearest neighbors' algorithm
ANN	Artificial Neural Network
DNFS	Deep Neural Fuzzy System
DT	Decision Tree
NB	Naive Bayes
ECG	Electrocardiogram
HF	Heart Failure
DCM	Dilated cardiomyopathy
OS	Operating System
WHO	World Health Organization
SMO	Sequential Minimal Optimization
WEKA	Waikato Environment for Knowledge Analysis
SSADM	Structured Systems Analysis and Design Methodology
OOD	Object Oriented design
OOP	Object Oriented Programming
UI	User Interface
TIFF	Tagged Image File Format
BMP	Basic Metabolic Panel
PNG	Portable Graphics Format

Chapter 1 - Introduction

1.1 Chapter Overview

This chapter is focused on the heart failure prediction system for heart patients, cardiologists and medical specialists to diagnose whether the patient is suffering from heart failure or not. This chapter covers the problem background, definition, solution, aim, research gap, research questions, project scope, requirements, and objectives in our project.

1.2 Problem Background

Heart failure occurs when the heart muscles are unable to supply adequate blood flow around the body due to certain heart conditions such as narrowed arteries in the heart or high blood pressure, gradually leaving the heart too weak or stiff to fill and pump blood properly (Lona and Mathue, 2021). When this happens, blood often backs up and fluid can build up in the lungs, causing shortness of breath. At first, the heart stretches to contract more strongly and keep up with the demand to pump more blood (Mayo *et al.*, 2021). Over time this happens, causing the heart to become enlarged. Due to this contraction, these contracting cells of the heart get bigger and increase the muscle mass also. Then the body tries to compensate in other ways. As an example, the blood vessels narrow to keep blood pressure up, trying to make up for the heart's loss of power. These temporary measures mask the problem of heart failure, but they don't solve it (Loana Dumitru *et al.*, 2021). Heart failure continues and worsens until these compensating processes no longer work. Heart failure can involve the heart's left side, right side or both sides (Malik A, Brito D, and Chhabra L., 2020). But it usually affects the left side of the heart first. There are four stages of heart failure called stage A, stage B, stage C and stage D which range from high risk of developing heart failure to advanced heart failure (Malik A, Brito D, and Chhabra L., 2020).

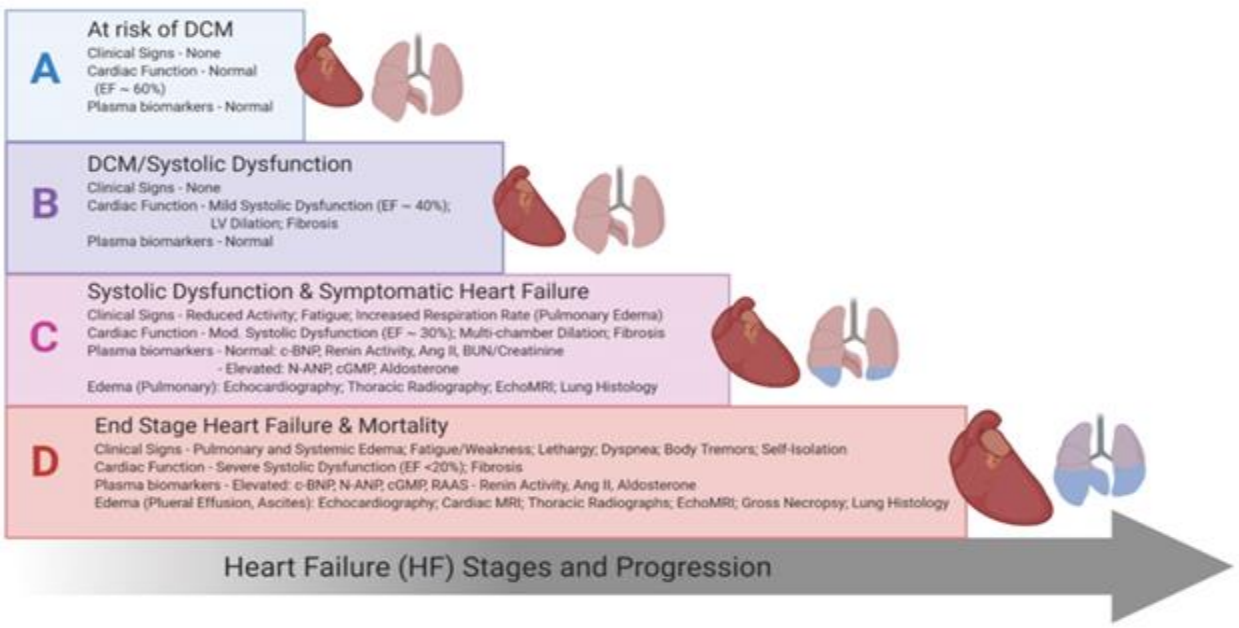


Figure 1. 1: Different Stages of Heart Failure (Tripathi, Sullivan, Fan and Mehta, 2021)

Receiving the proper treatment at the right stage during heart failure is very important to lower the progression of heart failure (Malik A, Brito D, and Chhabra L., 2020). If you have recently been diagnosed with heart failure, your treatment plan will depend on the stage at which your heart failure has been determined. Treatment of heart failure depends on the underlying cause, and this will direct the main treatment to prevent further deterioration (Mayo *et al.*, 2021). Treatment for heart failure usually aims to control the symptoms for as long as possible and slow down the progression of the condition. A careful history of the patient is very important for the diagnosis and to identify the cause of heart failure. Family history of the heart diseases, smoking status, hyperlipidemia, hypertension, and diabetes mellitus are some of the factors that should be taken into account during the assessment of the patient history in order to draw a risk profile of the patient (Mayo *et al.*, 2021). Whether anyone has any symptoms most people don't know what will happen and will not care about them until the last moment because of the lack of expertise or Sometimes patients are unable to access healthcare services due to some reasons. This makes heart disease a major concern to be dealt with. It is very important to diagnose heart failure at the early stages (Tripathi, Sullivan, Fan and Mehta, 2021). Nowadays, there are many kinds of heart

diseases, which consist of their own set of symptoms and treatment options. The below table shows some common types of heart diseases.

Heart Disease	**Description**
High Blood Pressure	The pumping rate of the blood against the artery walls is very high.
Stroke	The blood supply to the part of the brain is interrupted or decreased.
Congestive heart failure	The heart isn't pumping blood as well as it should be.
Congenital heart disease	It is a heart abnormality that develops before birth.
Cardiac arrest disease	Reduces the delivery of oxygenated blood to the cardiac muscle.
Coronary artery disease	Vascular injuries occur due to cholesterol layers buildup in the arteries.
Peripheral artery disease	The veins carrying blood from the heart to the legs constrict or get blocked.
Arrhythmia	Problem of the rate or rhythm of the heartbeat.

Table 1. 1: Different Types of Heart Diseases (Marimuthu et al., 2018)

The major challenge in heart disease is its detection. There are several instruments available nowadays to predict heart diseases, but the issue is they are very expensive or less efficient. Early detection of heart diseases can decrease the mortality rate and overall complications. However, it is not possible to monitor patients every day in all cases accurately and consultation of a patient for 24 hours by a doctor is not available since it requires more sapience, time, and expertise.

Feature	Our Product	Healthcare+	Cardiio	PulsePoint
Multi-Platform Support	✔	-	-	✔
Heart Disease Prediction	✔	✔	-	-
ECG Reading	✔	-	-	-
Alerts	✔	✔	✔	✔
Community	+	+	++	+++
Multi Language Support	-	-	✔	✔
Patient Library Maintenance	✔	✔	✔	✔
Cost	$	$	$$$	$

Table 1. 2: Feature Comparison Chart.

1.3 Problem Statement

The majority of patients have no idea if they have heart disease or not, and if they have a heart disease what is the current stage of the risk level.

1.4 Research Gap

This research mainly focuses on building a system that can identify, predict cardiovascular diseases. The system can discover and extract hidden knowledge associated with diseases from a historical heart data set. Heart disease prediction system aims to exploit data mining techniques on medical data sets to assist in the prediction of heart diseases.

1.5 Research Questions

- How can previously recorded data related to heart disease be helpful to find out or predict another person's heart disease?
- How to extract data from an ECG soft copy using image-processing?
- By what means would we be able to transform information into helpful data that can empower medicinal services practitioners to settle on viable clinical decisions?

1.6 Research Aim

One of the main goals of this system is to provide a tool for doctors and patients to detect heart disease at an early stage when entering the data about the patient or ECG by using this system.

This in turn will help to provide effective treatment to patients according to their current stage and avoid severe consequences. When the user enters the input values from the patient's health report, data will be fed into a model which predicts the probability of having heart disease. If the patient suffers from heart failure, the system will detect the patient’s current heart failure stage and provide some suggestions according to their present condition, and suggest some treatments as well. The system will recommend some cardiologists and specialized doctors according to the patient's current condition. System also provides the contact information of various cardiologists and medical specialists with whom the patients can easily connect via the platform.

1.7 Project Scope

1.7.1 In - Scope

- Collecting details of every user for the registration.
- Predicting if the user has a chance of being a heart patient.
- Categorizing the stages of the disease and giving instructions for the stages.
- ECG soft copies can be scanned.
- ECG soft copies must be acquired with a resolution of 300 or 600 DPI',either in gray-scale or color and stored in PNG,TIFF,BMP or JPG formats.

1.7.2 Out - Scope

- Application is not fully automated; the user must insert most of the data manually in the beginning.
- Cannot scan the ECG because it is too long to get pictures of the copy.
- This product will only support the English language.

1.8 Rich Picture Diagram

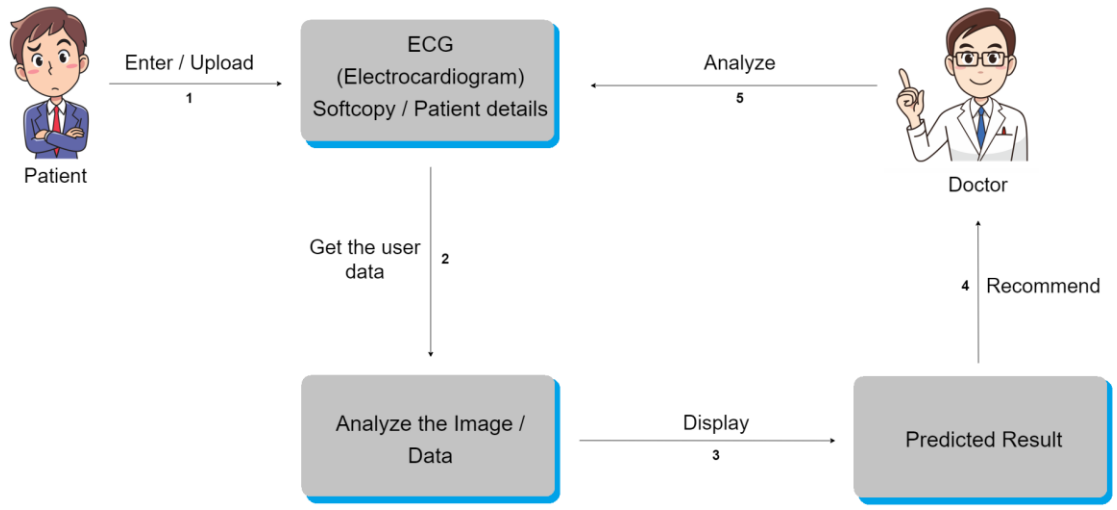


Figure 1. 2: Rich Picture Diagram

Rich picture is explained more below

1) Patient should enter the required data / softcopy of ECG to the system to do the prediction.
2) The system will get the data and analyze them accordingly to do the prediction.
3) After the prediction will display the predicted result to the patient.
4) If the prediction decides the patient has a heart disease system will recommend a doctor.
5) The doctors can go through the patient details for further medications.

1.9. Objectives

1.9.1 Research Objectives

- Discovering the uses of data science and machine learning techniques in a healthcare setting.
- Learning about Image Processing.

1.9.2 Academic Objectives

- Improve Research and Analytical Skills
- Improve Specification and Comparison Skills
- Get an Understanding of various Technologies and Techniques used in machine learning and image processing.
- Improving time management skills along with communication skills.

1.9.3 Operational Objectives

- Research Data Science and Machine Learning Trends.
- Research and identify the weak areas in existing solutions.
- Research about algorithms and datasets.
- Identify the software and hardware requirements.
- Identify the room to grow in the future.
- Create a simple, easy to understand and a user-friendly application.

1.10 Resource Requirements

Hardware Requirements	Software Requirements
Mobile Device Running Android OS	Windows OS
Laptop Computer	Android OS
	Jupyter Notebook
	IntelliJ
	Adobe XD
	PyCharm

Table 1. 3: Hardware Requirements and Software Requirements

Technology Stack

React Native	Front-end development
Python	Back-end development
GitHub	Version controller
Click Up	Work management application
Slack	Communication Platform
Google drive	Storage solution
Google Docs	Collaboration medium for writing

Table 1. 4: Technology Stack

1.10.3 Data Requirements

Heart Diseases Dataset

The data set called “Heart Diseases” from University of California Irvine consists of 4 databases of heart disease patients from Cleveland, Hungary, Switzerland, and the VA Long Beach. This dataset mainly consists of patient’s,

- Age
- Gender
- Chest Pain Type (#cp)
  -- Value 1: Typical Angina
  -- Value 2: Atypical Angina
  -- Value 3: Non-Anginal Pain
  -- Value 4: Asymptomatic
- Resting Blood Pressure(#trestbps)
- Serum Cholesterol in mg/dl (#chol)
- Fasting Blood Sugar (#fbs)
  -- fbs > 120 mg/dl (1 = true; 0 = false)
- Resting Electrocardiographic Results (#restecg)
- Maximum Heart Rate Achieved(#thalach)
- Exercise Induced Angina (#exang)
  -- 1 = yes; 0 = no
- ST Depression Induced by Exercise Relative to Rest (#oldpeak)
- The Slope of the Peak Exercise ST Segment(#slope)
  -- Value 1: upsloping
  -- Value 2: flat
  -- Value 3: down sloping
- Number of Major Vessels (0-3) Colored by Fluoroscopy (#ca)
  3 = Normal; 6 = Fixed Defect; 7 = Reversible Defect (#thal)
- Diagnosis of Heart Disease (angiographic disease status) (#num) - **The Predicted Attribute.**

-- Value 0: < 50% diameter narrowing

-- Value 1: > 50% diameter narrowing

Electrocardiography Images Dataset

ECG images dataset of Cardiac Patients created under the auspices of Ch. Pervaiz Elahi Institute of Cardiology Multan, Pakistan that aims to help the scientific community for conducting the research for cardiovascular diseases. This dataset consists of ECG images of,

- Myocardial infarction patients
- Patients that have abnormal heartbeats
- Patients that have history of Myocardial infarction
- COVID-19 patients
- Normal people.

1.11 Chapter Summary

This chapter provided an overall introduction to the proposed project. Here explained what heart diseases are, the necessity of a handheld mobile device to analyze health and well being, what is the aim, the technologies expected to use in the development and implementation phase.

Chapter 2 - Literature Review

2.1 Chapter Introduction

In the previous chapter introduced the cardiovascular disease application. In this Literature Review chapter, will dive into previous studies carried out related to this project. Here, identify the approaches the previous researchers have used and evaluate the machine learning algorithms and image processing algorithms they have used.

2.2 Previous Proposed Solution Overview

According to the World Health Organization (WHO), cardiovascular diseases are a primary cause of death for approximately around 17.9 million people in each year. A.T. Matthias and his team conducted research on patients' physical activity levels before acute coronary syndrome using their experience at a tertiary care hospital in Sri Lanka, and the results are presented below (Matthias, Indrakumar and Gunatilake, 2018).

"The study included 348 males and 156 females. The majority 393 (77%) of the patients were between 45 and 74. Pre-existing ischemic heart disease was found in 236 (46.8%) patients, diabetes in 198 (39.3%), hypertension in 254 (50.4%) and dyslipidemia in 109 (21.6%). 123 (24.4%) were current smokers, 107 (21.23%) were former smokers and 274 (54.36%) were non-smokers.The chief complaint on admission was chest pain in 347 (68.8%) and shortness of breath in 78 (15.4%). Of the total of 504patients, 128 (25.1%) were highly active, 87 (17.1%) were moderately active and 289 (56.7%) belonged to the low activitygroup. When considering METS per week 134 (26.3%) patients spent less than 1000 METS/week." (Matthias, Indrakumar and Gunatilake, 2018)

Considering those risk factors and results, the team decided to help patients by developing an app that can predict a person's risk of heart disease based on the information provided by the patient. Whether there are products that are planned to exist, try to make full efforts in order to make the

results better and more trustworthy because this is a more sensitive matter which some lives depend on our application. Doctors can be contacted via the app for treatments and further clarifications.

2.3 Existing Work

2.3.1 Various Existing Works and Results

Various works have been done related to heart disease prediction using machine learning algorithms and data mining techniques.

K. Polaraju et al proposed a model for predicting heart diseases by using Multiple Regression Model. It proved that Multiple Linear Regression is suitable for predicting heart diseases (Bavani, Josephine and Prasannakumari, 2019). The dataset which is used for the research is divided into two main parts. 70% data used for the training and 30% used for the testing purposes (Singh, Singh and Pandi-Jain, 2018). Based on those results they identified that the Regression algorithm is most appropriate when compared to other algorithms (Loana Dumitru *et al.*, 2021). Marjia et al developed a system for heart disease prediction using Bayes Net, KStar, j48 and Multilayer perception using WEKA software. But they came up with a situation where the level of accuracy is very low (Arul Jothi, Subburam, Umadevi and Hemavathy, 2021). Different kinds of algorithms recommended by Ashok Kumar Dwivedi et al such as Logistic Regression, Classification Tree, and Naive Bayes. Among all these algorithms, Logistic Regression gave best accuracy compared to other algorithms (Marimuthu et al., 2018).

A method which predicts heart disease using a patient's medical dataset was recommended by Ashwini shetty et al. To build the system, there are 13 risk factors attributes used as input attributes. Data integration and data cleaning was performed after analysing the data from the dataset (Singh, Singh and Pandi-Jain, 2018). Artificial neural network was recommended by Noura Ajam to diagnos heart diseases. Feed forward Backpropagation learning algorithms have been used for testing the model. Results showed the 88% accuracy level and twenty neurons in the hidden layer (Arul Jothi, Subburam, Umadevi and Hemavathy, 2021). Prajakta Ghadge et al, have suggested big data for prediction of heart attack. The main objective of this is to provide a prototype using data modelling techniques and big data. It is also used for extracting the patterns

and relationships from the database which are associated with heart disease (Mayo et al., 2021). This system mainly consists of 2 databases which are, original big dataset and updated one. This system can assist healthcare practitioners to make intelligent decisions. But this system is not fully automated. S.Prabhavathi et al, have proposed a decision tree based Neural Fuzzy System (DNFS) technique to analyse and predict various heart diseases (Bavani, Rajini, Josephine and Prasannakumari, 2019). Different data mining techniques have been used to enhance heart disease prediction. The results showed that Support vector machines (SVMs) and neural networks were highly recommended for predicting heart diseases. But the issue is data mining techniques not encouraging for heart disease prediction (Malik A, Brito D, and Chhabra L., 2020).

To predict heart diseases, k-means and naïve bayes are used by Sairabi H.Mujawar et al. This system was built using the historical heart database that gives diagnosis (Singh, Singh and Pandi-Jain, 2018). There are 13 attributes that have been considered for building the system (Arul Jothi, Subburam, Umadevi and Hemavathy, 2021). This system is used to predict whether the patient has heart disease or not based on the values of these 13 attributes (Bavani, Rajini, Josephine and Prasannakumari, 2019). To predict the heart disease using a smaller number of attributes Sharan Monica.L et al proposed a method using data mining techniques. Predict the risk level of the heart disease of a patient, proposed a method by Sharma Purushottam et al. Partial tree technique and c45 rules have been used for this method. According to the results, it showed that the potential prediction is more efficient (Marimuthu et al., 2018).

Below table shows the Comparative analysis between various existing works with used techniques and their accuracy levels.

YEAR	AUTHOR	PURPOSE	USED TECHNIQUES	ACCURACY LEVEL (%)
2014	Liu et al.	Prediction of Heart Disease using Support Vector Machine Algorithm and K-nearest neighbors algorithm	Support Vector Machine Algorithm (SVM)	95%
			K-nearest neighbors algorithm (KNN)	91.49%
2015	Sharma Purushottam et al.	Efficient OF Heart Disease Prediction	Decision tree classifier	86.3%
2015	Sairabi H. Mujawar et al.	Prediction of Heart Disease	Modified k-means algorithm Naive Bayes algorithm.	93%
2015	Aljaaf et al.	Multi level risk assessment of developing HF(Five risk levels of HF)	Decision tree classifier	86.30%
2015	Noura Ajam et al.	Heart Disease Diagnoses	Artificial Neural Network(ANN)	88%

2016	Marjia et al.	Prediction of Heart Disease using the WEKA tool.	K Star	75%
			Bayes Net	87%
			Multilayer Perceptron	86%
2016	Ashok Kumar Dwivedi et al.	Evaluate the performance of different machine learning techniques for heart disease prediction.	Naive Bayes	83%
			K-nearest neighbors algorithm (KNN)	80%
			Logistic Regression	85%
			Classification Tree	77%
2016	Jayamin Patel et al.	Heart Disease Prediction using Machine Learning and Data Mining Technique	J48	56.76%
			Logistic model tree	55.75%
2016	Ashwini Shetty A et al.	Different Data Mining Approaches for Predicting Heart Disease	WEKA tool, MATLAB. Neural Network	84%
			Hybrid Systems	89%

2018	R. Sharmila et al.	A conceptual method to enhance the prediction of heart diseases using the data techniques.	Support Vector Machine Algorithm (SVM) in parallel fashion	85%

Table 2. 1 Comparative analysis between various existing works with used techniques and their accuracy levels (Almustafa, 2020)

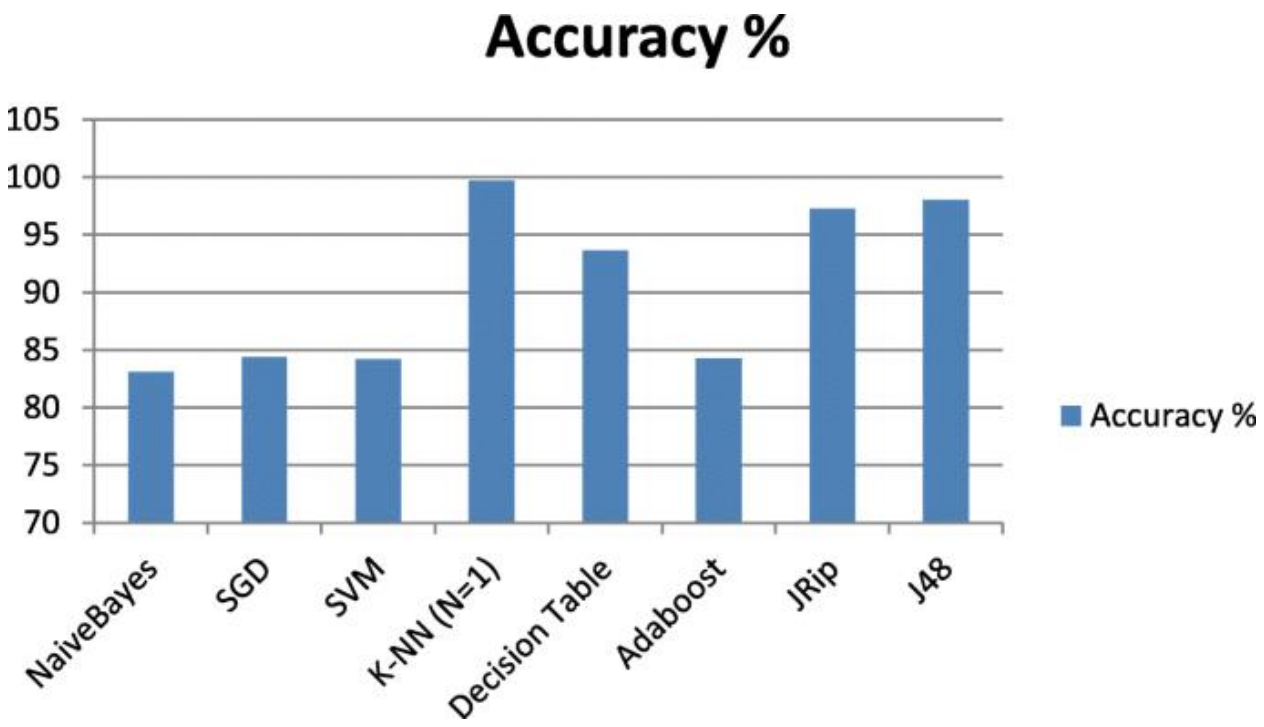


Figure 2. 1: Various Machine learning techniques with their accuracy levels (Almustafa, 2020)

2.3.1 Reviews on Existing Research

K. Polaraju developed a prediction of heart disease using multiple regression models.It demonstrates that multiple linear regression is acceptable for predicting heart failure (Arul Jothi, Subburam, Umadevi and Hemavathy, 2021). It was made using a specified data set which consists of 3000 cases with 13 distinct properties. The data set is separated into two sections as testing and training (Bavani, Rajini, Josephine and Prasannakumari, 2019). Based on data, getting a clear understanding of the regression method leads to the other algorithms in terms of classification accuracy (Singh and Pandi-Jain, 2018). Marija produced a cardiovascular prediction system with the help of j48, SMO, KStar, and Bayes Net, as well as Multilayer perceptron using WEKA software. Using k-fold cross validation SMO and Bayes Net outperform KStar, Multilayer perceptron, and J48 approaches in terms of performance from multiple factors (Marimuthu et al., 2018). The efficiency of the performance of algorithms is limited. As a result, the validity of the performance is raised to provide better solutions for this cardiovascular diagnosis. Marjia designed a heart disease prediction model using KStar, j48, SMO, and Bayes Net as well as Multilayer sensory using WEKA software (Arul Jothi, Subburam, Umadevi and Hemavathy, 2021). The accuracy performances of the algorithms are still unsatisfactory. As a result, the accuracy of the performance is increased further to provide better decisions for illness diagnosis (Bavani, Rajini, Josephine and Prasannakumari, 2019). S.Seema focuses on strategies for predicting chronic disease by mining data from past health records utilizing Naive Bayes, Decision Trees, Support Vector Machines (SVM), and Artificial Neural Networks (ANN). Comparison research is conducted on classifiers to see which performs better on an accuracy rate. In this experiment, SVM has the best accuracy rate, but for diabetes, Naive Bayes has the highest accuracy (Singh, Singh and Pandi-Jain, 2018).

2.4 Tools and techniques

When it comes to machine learning there are data analysis algorithms used to continuously learn from vast amounts of data. As models are exposed to new sets of data, they adapt to produce consistent and reliable output. There are various tools that are used in machine learning and they have various reasons why they are important. Learning about different tools and choosing the correct tools will help improve product reliability and accuracy.

2.4.1 Review of Tools & Techniques

- **K-nearest-neighbor (KNN)**

  K-nearest-neighbor is a very simple and fundamental classification technique which can be used when the data distribution is unknown. It is usually performed when uncertainties in the probability densities are not significant. The k-nearest-neighbor classifier is mainly based on the Euclidean distance value. Euclidean distance is the value between a test sample and the specified training samples.

  $$d(x_i, x_l) = \sqrt{[(x_i1 - x_l1)^2 + (x_i2 - x_l2)^2 + \cdots + (x_ip - x_lp)^2]}$$

  The K of the K-nearest neighbor represents the number of nearest values which compare with the given input to the system (Peterson, 2009).
  Using the KNN algorithm will help the system to predict the output result more correctly and reliably by analyzing and calculating the pre-existing dataset with the user input.

- **Decision Trees (DT)**

  The Decision Tree (DT) is a supervised learning strategy for solving classification problems. It is a classifier in the structure of a tree. Decision Trees are easy to understand since they are similar to how humans think while making a decision (Machine Learning Decision Tree Classification Algorithm - Javatpoint, no date). Using decision trees will make it easier to build and train the system. So, the final result will be more understandable and accurate based on the dataset and the model.

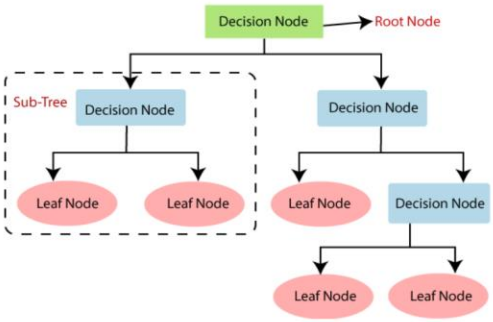


Figure 2. 2: example decision tree (Machine Learning Decision Tree Classification Algorithm - Java point, no date)

- **Genetic Algorithm (GA)**

  Genetic Algorithms are mostly inspired by the natural selection of genetics. This algorithm uses the intelligent exploitation of random historical data to find a better solution. These are mostly used to generate high quality solutions for the problems. Genetic Algorithms simulate a process of natural selection in order to find the species which can adapt to their environment perfectly. The best Species will be selected for the next generation and improved again. (Corus *et al.*, 2018)

- **Naive Bayes (NB)**

  Naïve Bayes is one of the most popular machine learning and data mining techniques. Its efficiency comes from the assumption of attribute independence. This helps to classify the data based on the conditional probability values computation yet at the same time maintains simplicity and efficiency. This algorithm is fast and is mostly used in real time processes. Naive Bayes is easy to learn, more efficient than other algorithms that can be useful when it comes to spam direction. (Chen et al., 2020)

2.4.2 Review of evaluation Techniques

Technique	Advantages	Disadvantages
K-nearest-neighbor (KNN)	● Doesn't need a training period to train the functionalities. ● Adding new data is an easy task.	● Doesn't work well with high dimensional datasets because It becomes difficult for the algorithm when calculating the distance of each dimension.
Decision Trees (DT)	● Decision Trees are like human thinking while making a decision, so it is easy to understand. ● Domain specific data can be acquired more easily. ● Helps to think about all the possible outcomes for the problem.	● A small change in data can cause instabilities in the tree structure. ● Takes a higher time to train the model. ● The calculations can be complex.
Genetic Algorithms (GA)	● Always generates a solution that improves with time or iterations. ● Faster & efficient than other classical algorithms.	● Sometimes cannot guarantee the solution provided by the algorithm. ● Must be careful in implementation or else the system will fail to reach the optimal solution. ● The fitness value can be increased after generations, so the computation gets more & more complex.

Naive Bayes (NB)	● When it comes to making assumptions naive bayes performed better compared to other algorithms. ● Easy to implement & takes only a small training time.	● When it comes to assumptions of independent predictors in real life it is almost impossible to get independent predictors but naive bayes use it.

Table 2. 2: Advantages and Disadvantages of used techniques (Almustafa, 2020)

2.5 Selected Approach

Doing this research on various technologies & ideas helped to decide what will be the best for the research's problem domain. With the knowledge that the team gained while doing the literature so far, the team decided to build the core of our prediction system with a key algorithm that was discussed above. K-nearest-neighbor (KNN) is the machine learning algorithm the team decided to use along with the project's core. With the extracted data and efficiency of the KNN algorithm, hope the team could achieve greater results with a good accuracy level when predicting heart disease.

2.6 Chapter Summary

In this chapter discussed about the literature review of our project. So, here explained the various heart disease prediction systems done by different people using data mining and machine learning techniques and what are the results of those existing works. And also what are the techniques and methods they used for and accuracy levels of those techniques also explained in this chapter.

Chapter 3 - Methodology

3.1 Chapter Overview

In the previous chapter, focused on the Literature review related to our project. In this Methodology chapter, mainly discuss the whole development of the software with the methodologies. The risk and mitigations, Gantt chart diagram, Team Work Breakdown Structure (WBS), usage of project management, etc. will also be discussed in this chapter. The requirements for this are clearly identified and the outcome of the project is also clear. Here the most appropriate methodologies for the development of the software with the team members are discussed.

3.2 Research Methodology

Selecting & using the correct type of methodology is an important part when it comes to a research problem. The correct use of methodologies helps to improve the reliability and validity of the research. There are two types of mainly used methodologies that can be helpful for any research project.

3.2.1 Types of Methodologies

Qualitative research

Qualitative research will help improve research that collects textual data and analyzes them. Like gathering feedback from certain groups of people or domain experts. This is best for describing or diving in depth knowledge in a specific concept. Qualitative research techniques are for text type data so mostly can’t fall into numbers. There are strategies to interpret and reveal concepts and meanings to be used with qualitative research techniques. (Eshlaghy et al., 2011)

Quantitative research

Quantitative research would be a great solution for researchers that used statistics & some types of data mostly numerical to measure & test in a problem. This methodology helps to aim and produce generalisable knowledge about the causes of a problem. To build a valid research project requires carefully gathered data that can be used by other researchers. There are three broad

classifications of quantitative research which are: descriptive, experimental, and causal comparative. (Abood and Alalwany, 2021)

3.2.2 Selecting the Research Methodology

Since our research focuses on heart complication detecting systems it requires data from users mostly numerical or an image that later extracted to numerical data from. So mostly in this research project we should use numerical data to work with & produce the output so going with the Quantitative research methodology would be appropriate to improve our research reliability & validity.

3.3 Development Methodology

Successful software development projects are always managed well. The development team should choose the best software development methodology for the project in order to manage the project efficiently. All the methodologies have different advantages and disadvantages for different reasons. Now let's see the most used development methodologies.

3.6.1 Types of Development Methodologies

- **Waterfall**

  Waterfall model is a linear, sequential approach which divides software development into pre-defined phases. Each phase must be fully completed before the next phase begins. So, there is no overlapping between any phases. And each phase has specific deliverables and a review process. This methodology is most suitable for small projects where requirements are well defined (Dearnley, 2021).

- **Spiral**

  Spiral model is a risk driven software development process model which is a combination of waterfall model and the iterative model. This model is most suitable for highly risky and complex large projects that might take a long time to be developed and delivered where the requirements are ambiguous. Additional changes can be done in later stages also (Chandra, 2015).

- **Agile**

  Agile methodology is a way to manage a project by breaking into several phases. It is based on an iterative and incremental approach. Here, both testing activities and development are concurrent (Saeedi and Visvizi, 2021).

- **Prototype**

  The prototyping model is a systems development process in which a prototype is produced, tested, and changed as needed until an acceptable result is reached from which the entire system or product may be developed. This model is an expensive model that is used for building a product, testing a concept or learning purposes with unclear requirements (Dearnley, 2021).

3.6.2 Selecting the Development Methodology

Among all these development methodologies, Agile Methodology is selected for this project. Doctors and patients use this system and sometimes, they want some new requirements. So agile methodology has flexibility, managing uncertainty, continuous improvement, and speed. Since this is a small project and some models such as spiral, prototype model will not work well for this project. Therefore, the team like to go with the Agile method. The reasons for not using other models for this project are shown in the table below.

Model	Reason for not choosing the model
Waterfall	It does not allow much reflection or revision
Spiral	Does not work well for small projects Risk analysis is involved in every cycle of development.
Prototype	It is not suitable for small projects as it is very expensive. It is more complex than other models.
Rapid Application Development Model	Skills and a good technical knowledge is required.

Table 3. 1: Reasons for not choosing the other models

3.4 Design methodology

Choosing the suitable design methodology is an important decision in a research problem. The correct use of methodologies improves the structural integrity of any program that is planned to create. There are two types of methodologies that can be helpful in software development projects.

3.4.1 Types of Design Methodologies

Object Oriented design (OOD)

Object Oriented Design allows programmers to design patterns that incarnate expert knowledge distilled from the practical experience in object-oriented design, in a compact and reusable form. Object Oriented Design is a practice of creating a software architecture that enables flexibility through modular design (Smith, 2011). OOD is an approach to analyze & design a software system by using Object Oriented concepts. Object Oriented Design gives the programmers the possibility of creating relationships between identified objects (Bontchev and Milanova, 2020).

- Programs are easier to test & maintain when using OOD.
- Serves as part of the object-oriented programming process or lifecycle.
- This is easy to understand.

Structured systems analysis and design methodology (SSADM)

Structured systems analysis and design methodology is also known as Data Flow Diagrams. This uses a formal methodical approach to the analysis & design of a project. SSADM is a methodology based on the waterfall software development model. Main objective for the SSADM development methodology is to design a highly structured method which will deliver designs for workable, maintainable, extensible and portable computer systems (Rose, 1991).

- SSADM offers better understanding.
- This improves the useability.

3.4.2 Selecting the Design Methodology

After analyzing and knowing all the advantages of either above mentioned design methodologies the team decided to choose Object Oriented Design (OOD) as the design methodology for this project. With the use of Object-Oriented Design (OOD) the team can gain the benefits and learn the concepts of Object-Oriented Programming (OOP) such as abstraction, encapsulation, polymorphism, and inheritance.

3.5 Evaluation methodology

Evaluation is the process of finding the usefulness & how successful is the project. At first, we gathered the users' opinions by sharing an open questionnaire created in google forms among the people that can be the future users of the product intended to be created within this project. So, we could evaluate how useful this project can be to the end users. All the data we gathered through the questionnaire will be broadly discussed in chapter four. To evaluate the successfulness of the product we build we plan to have meetings while the development process and brainstorm among the developers of the project what can be improved & allow some few selected users to use a

prototype of the product before publishing. So we can make the product more successful and useful.

3.6 Project Management Methodology

A project management methodology is a collection of practices and principles which will guide the process of structuring the project. So that it helps to manage the project in the best way possible. Among all project management methodologies, we are going to use agile project management methodology for our project. In Agile project management methodology, divides a project into several stages to make it easier to manage. It should involve constant collaboration with stakeholders as well as continual development at all stages. The process consists of several phases as you can see in the below figure.

Figure 3. 1: Different Phases of Agile Project Management Methodology (Kanban Versus Scrum in Agile Software Development, 2022)

Both with team members and project stakeholders, continuous cooperation is essential. Agile project management concepts have worked more iteratively and flexibly, allowing them to adapt to the project's shifting requirements as well as the delivery faster. Agile has most useful advantages which best for our project such as,

- Increased attention on the specific needs of user
- Increased satisfaction in users
- Increased flexibility allowing team to easily adapt to change
- Better project management
- Risk reduction
- Providing an opportunity for review and correction
- Increased frequency of feedback and collaboration
- Improved development process

3.7 Teamwork Breakdown Structure (WBS)

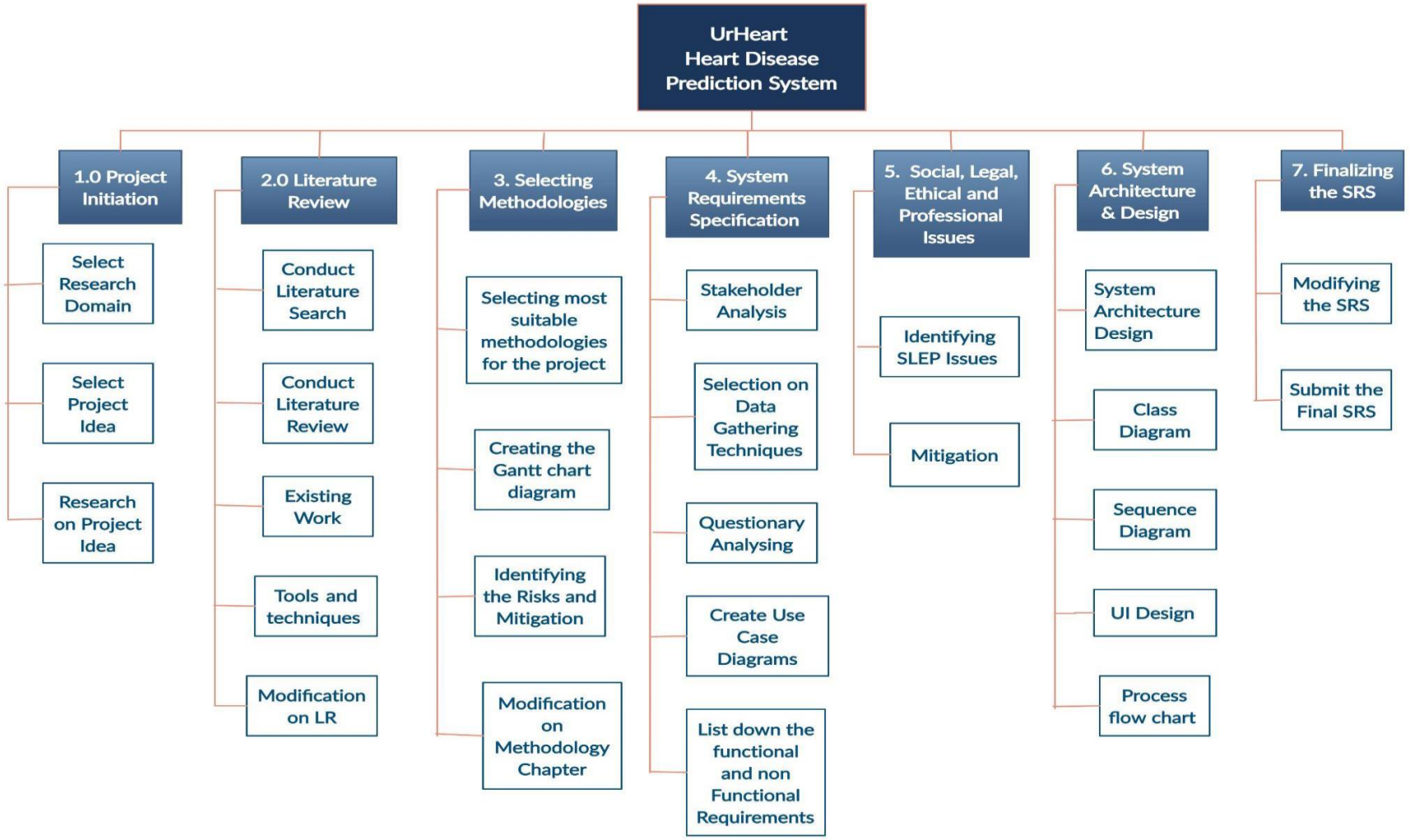


Figure 3. 2: Work Breakdown Structure

3.8 Gantt chart diagram

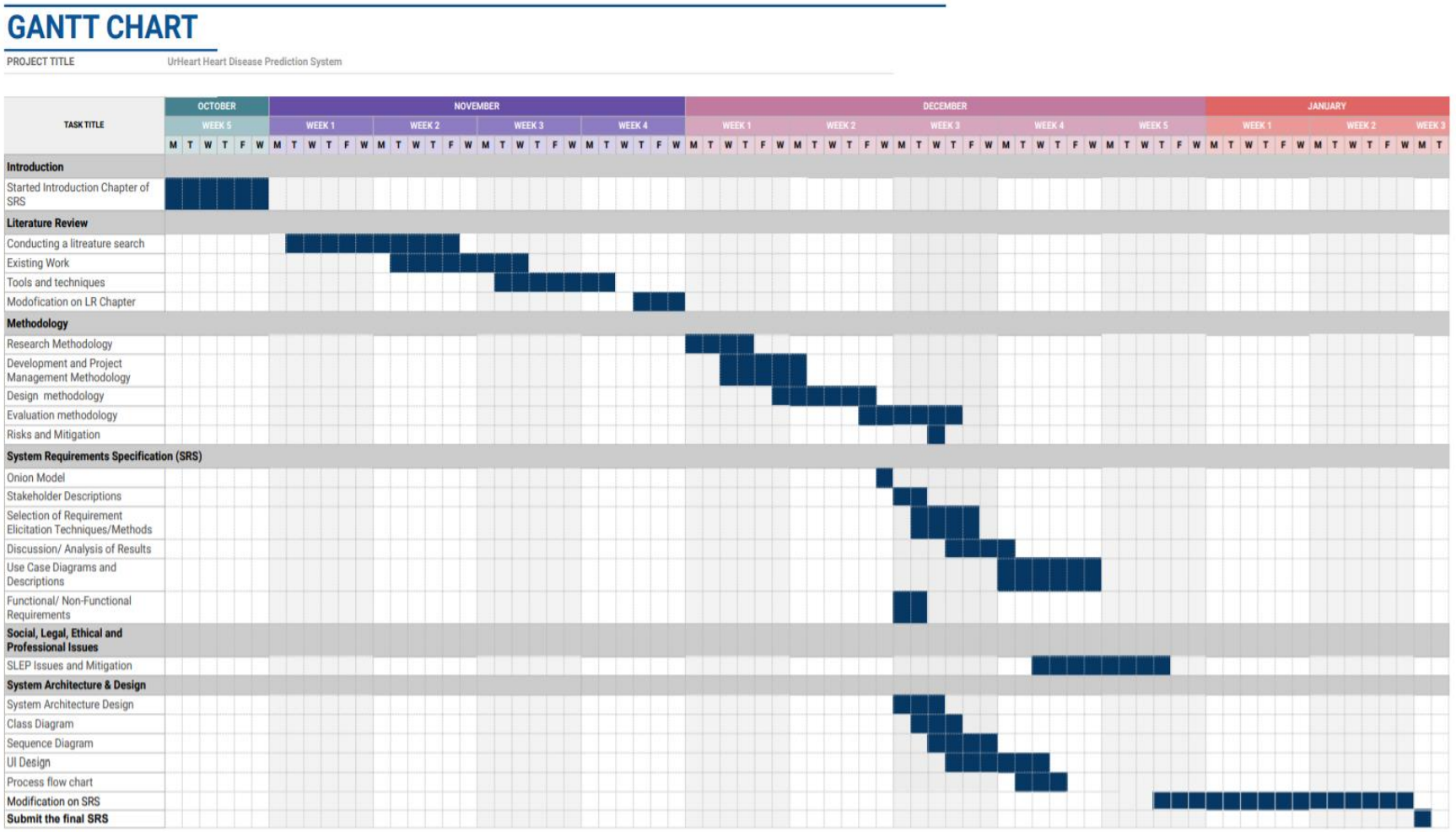


Figure 3. 3: Gantt chart diagram

3.9 Usage of Project Management and Collaboration Software In the project

Project management software are very useful because it allows us to keep track of activities and see them in context of the overall project. Also, it will be able to see how they fit into the wider picture and how they affect other activities that must be finished, are in the process of being completed or have already been completed. Also, it helps to complete client requirements and manage time, budget, and scope constraints. Used primarily as an application, project management software provides increased efficiency and project communication. In addition to that plan out the deliverables, prioritize tasks, high productivity, better flexibility, risk management and quality control are some usages of Project Management and Collaboration Software.

- **Google docs**

  To this project we used google docs to collaborate with the group members because the files can be edited and be saved at real time with no delays and can be edited from anware with the given access. It shows the history of the edits so if anything is missed we can restore to the previous versions quickly and easily. Can add comments & assign members to certain tasks easily.

- **Google meet**

  Throughout this entire project all the member virtual meetings were conducted through google meet because it's a wonderful medium that can be used by anyone and a trustworthy platform. We could share our thoughts, collaborate and brainstorm ideas among the team members using this software.

- **Google drive**

  To store & share the content among the team members we used google drive as the main solution. The free cloud space every user gets with every email account created in google was a great deal for every member throughout this project.

- **GitHub**

  GitHub is another popular & useful software used to collaborate & version control. This software has been used by many developers for the past years. We think this would be helpful to our team members to keep track of the product after the development process begins.

3.10 Risks and Mitigation

Risk Item	**Severity**	**Frequency**	**Mitigation Plan**
Having the low knowledge about technologies	High	High	Doing self-studies about the technologies with available resources and identifying the best and comfortable platform for our project by communicating with the group members.
Time management issues	High	High	Make a timetable and set personal deadlines for each task
Lack of accuracy of the information	Medium	Medium	Should work hard to make a proper database with valuable and very high accuracy level information and get that information from reliable resources and domain experts.
Technical failures and Bugs related to implementation	Medium	Medium	Should work hard and widen our knowledge on different types of technologies related to this project. And practicing them before applying to the project is a must.

Lack of responsibility	Medium	Medium	Doing the task which will be assigned in the proper manner and avoiding a lack of accountability and getting clear only works if everyone knows what it is they're working towards all of the time.
Lack of participation during group activities	Medium	Medium	Distribute work among the participants and everyone should be responsible for each task assigned.
Issues may be caused due to the health of the group members.	Low	Low	Plans work ahead

Table 3. 2: Risks and mitigation

3.11 Chapter Summary

This chapter consists of the methodologies and procedures used in this project. The risks involved as well as the mitigation plan for the system were identified. The most appropriate software development methodologies and project management methods were also discussed in detail. The activity schedule and project deliverables were also detailed. The work breakdown structure and Gantt chart diagram were also illustrated in this chapter.

Chapter 4 - System Requirements Specification (SRS)

4.1 Chapter Overview

This chapter will discuss System Requirements Specification and also dive into Stakeholder Analysis, Requirements Gathering, Functional and non-Functional Requirements.

4.2 Stakeholder Analysis

4.2.1 Onion Model

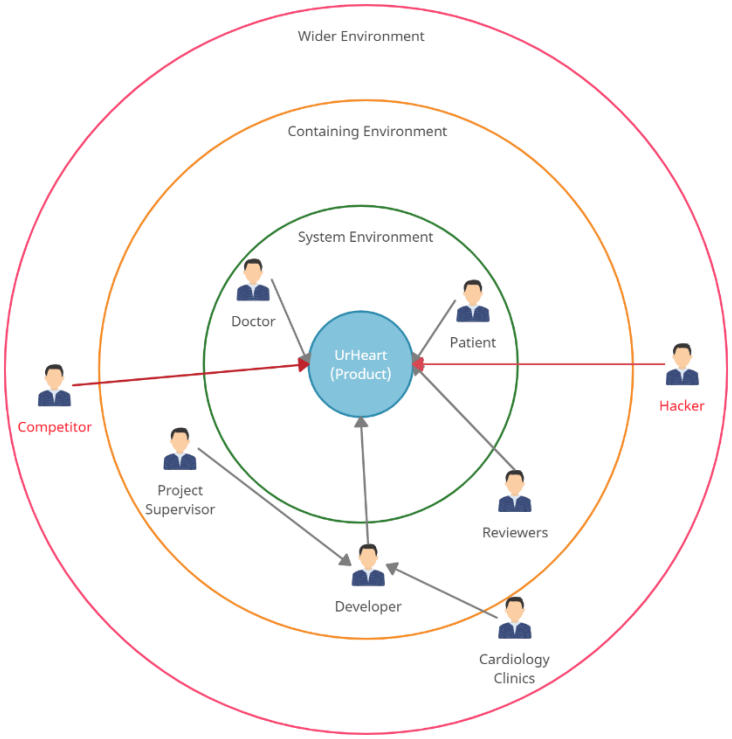


Figure 4. 1: Stakeholder Analysis Onion Model

4.2.2 Stakeholder Descriptions

Stakeholder	**Viewpoint**
Functional beneficiary	
Developers	Will improve functionalities and fix errors of the final products
Project Supervisor	Will help the developers create the best version of their product/ application
Cardiology Clinics	Will take the application output results and compare the results with their expertise and help make the application more accurate
Financial beneficiary	
Cardiology Clinics, Doctors	Making the predictions more accurate will call in more clients and application users, which will increase profits.
Social beneficiary	
Patients	The users will use this application to determine whether they are prone to a heart attack. If a patient/user was at risk, they can further contact a medical professional and start the healing process.
Operational beneficiary	
Developers	Come up with more app ideas, implement them and read user reviews, fix application errors, and shape the app to be more user friendly.
Negative Stakeholders	
Competitors	Wants to create a better application than ours.
Hackers	Wants to sabotage the functionalities of the system.

Regulatory	
Authorities	Want to regulate the user results and make sure the system works while not wanting the application to give out false information.
Experts	
Doctors, Cardiology Clinics	Help out developers with their questions and discuss their requirements extensively.
Neighboring systems	
Social Media	Make our product known among a wider range of audience.

Table 4. 1: Stakeholder Analysis Description

4.3 Selection of Requirement Elicitation Techniques/Methods

- Online Questionnaire

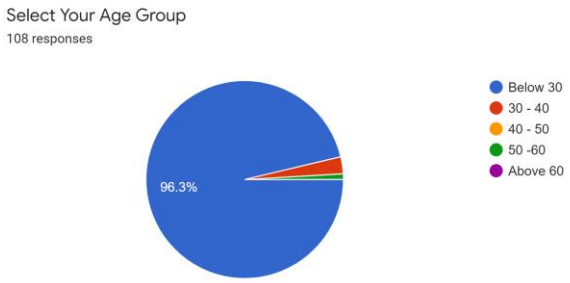
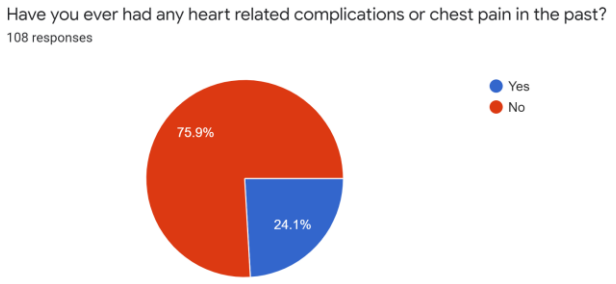
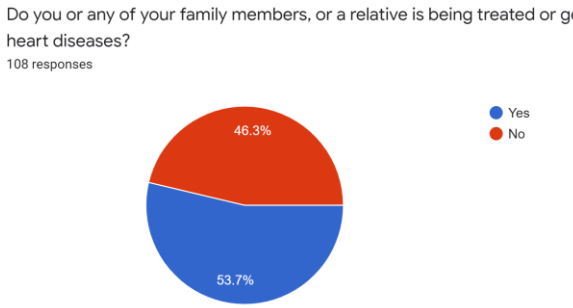
An online quiz was taken place to extensively collect data and identify the target audience. This was useful in getting a clear understanding about the users' preference.
Questionnaires should not be too long.

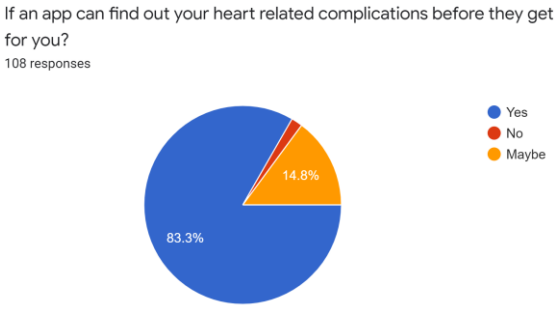
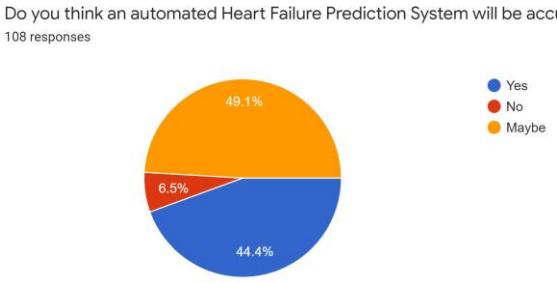
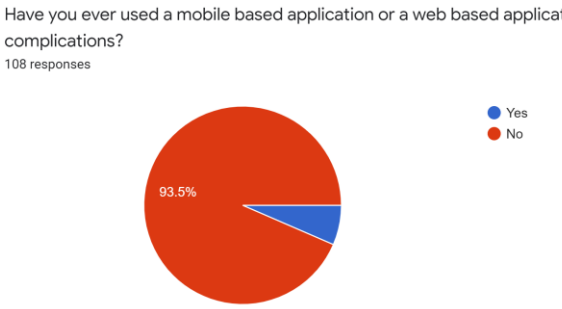
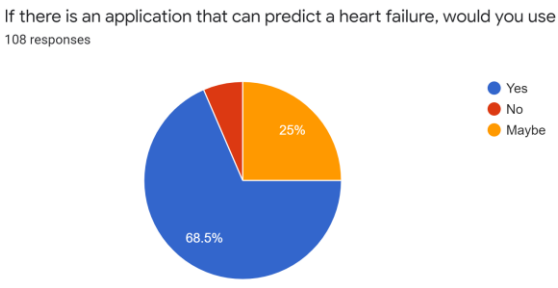
- Interviews

Interviews are the most capable and effective techniques accessible for data gathering. Anybody can learn to interview since interviews are simply a form of organized conversation between two or more individuals. To be effective in requirement gathering, the interviewer should come up with bright questions. As a requirement analyst, you must use interviews as a tool to get your point across to the users' and stakeholders' requirements and utilize this information to build a system that is well-suited.

4.4 Discussion/ Analysis of Results

An online survey was conducted to get a public insight on the issue. The survey was sent on 16th of November and the gathered answers are discussed below. 108 participants contributed to the above survey, most of whom are from the university.

	This question was placed to identify the age group of the participants who took place in the questionnaire. 96.3% of the participants were below the age of 30.
	Here 75.9% of the participants voted no, therefore it is safe to assume that the majority of the participants haven't experienced some sort of heart related complications from the given age group.
	But here when we asked, if anyone they knew is being treated for heart diseases, the majorities have been changed. The 53.7% of participants saying yes means, even though the participants have not experienced heart complications, they still have family members, friends or relatives being treated.

	Here 83.3% of the responders voted yes, further signifying the need for such an app in the market.
	49.1% of the participants aren't sure if an automated system would carry out accurate results but still 44.4% are confident that it would.
	Here 93.5% of the responders voted no, which means they aren't aware if such applications exist in the app store.
	Here 68.5% of the participants voted yes, meaning that the majority are willing to use such an application.

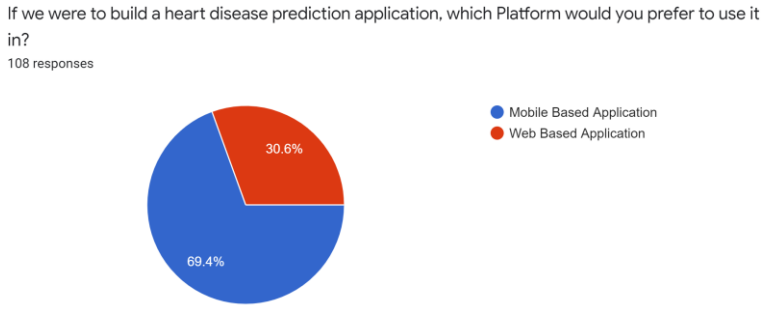
	Here 69.4% of the participants voted for a mobile based application rather than a web-based application, which means that most of the participants would prefer a mobile based application to a web-based application.

Table 4. 2: Discussion/Analysis Table

4.5 Use Case Diagram

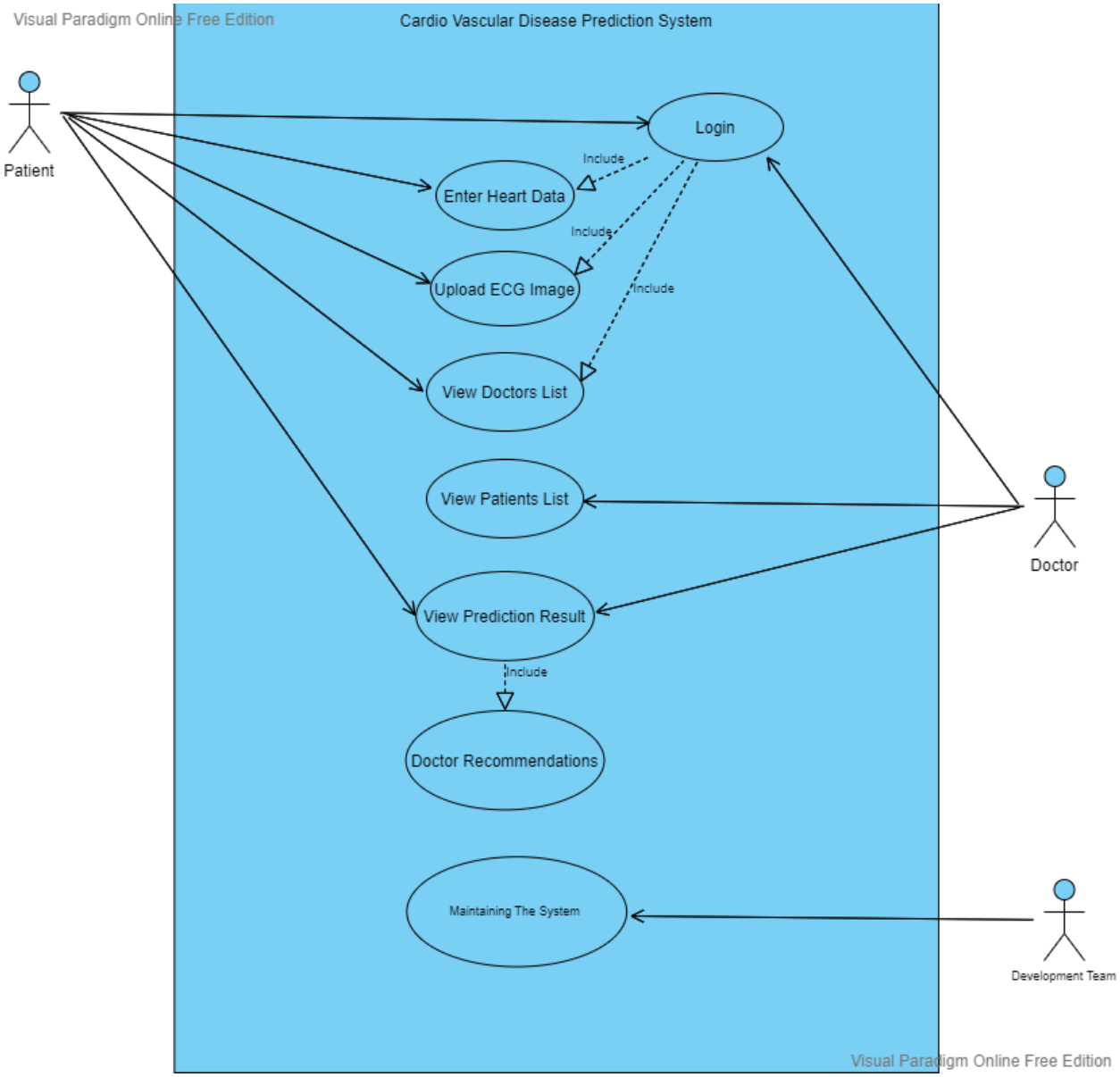


Figure 4. 2: Use Case Diagram

4.6 Use Case Descriptions

Use Case Name	Login	
Use Case ID	UC-001	
Description	The user is required to login to use the application features	
Priority	High-Level	
Primary Actor	Doctor, Patient	
Supporting Actors	None	
Pre-Conditions	User needs to be registered in the application	
Trigger	The user launches the application	
Main flow	**Actors**	**System**
	1) User opens the application 3) User chooses an option	2) Application displays options 4) Proceed with the option
Exception flow	**Actors**	**System**
	1) User opens the application 2) User enters incorrect login information.	3) Application displays an error message asking the user to check login information again.
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	Displays the home page	

Table 4. 3: Use Case Table 1

Use Case Name	Enter Heart Data	
Use Case ID	UC-002	
Description	User needs to enter heart details in order to use prediction data	
Priority	High-Level	
Primary Actor	Patient	
Supporting Actors	None	
Pre-Conditions	User needs to have the user profile updated	
Trigger	Displays the Heart Details page	
Main flow	**Actors**	**System**
	1) User selects "Predict Heart Disease" option 3) User selects values and updates the list.	2) Application displays the list of parameters that needs to be updated. 4) Values are recorded. 5) Application proceeds to process the information and outputs the prediction value.
Exception flow	**Actors**	**System**
	1) User selects "Predict Heart Disease" option 3) User misses out a parameter that needs to be filled.	2) Application displays the list of parameters that needs to be updated. 4) Application informs the user to fill all the parameters.
Alternate flow	**Actors**	**System**

Exclusions	None
Post Conditions	Information is processed.

Table 4. 4: Use Case Table 2

Use Case Name	Upload ECG	
Use Case ID	UC-003	
Description	User can upload an ECG image to get the prediction result	
Priority	High-Level	
Primary Actor	Patient	
Supporting Actors	None	
Pre-Conditions	The image must be confining to the given standards	
Trigger	Opens the user's mobile phone gallery.	
Main flow	**Actors**	**System**
	1) User selects “Upload ECG” option 3) User clicks on the Browse button and navigates through their mobile phone gallery to find the ECG Graph image.	2) Application displays Upload ECG page. 4) ECG graph is processed.
Exception flow	**Actors**	**System**
	1) User selects “Upload ECG” option 3) User clicks on the Browse button and navigates through	2) Application displays Upload ECG page. 4) ECG Graph image does not follow the requirements.

	their mobile phone gallery to find the ECG Graph image.	5) Displays an error message informing the user that the image does not confine to the required standards and asks the user to upload another image.
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	ECG graph image will be processed.	

Table 4. 5: Use Case Table 3

Use Case Name	View Doctors List	
Use Case ID	UC-004	
Description	List of doctors that can give the user feedback	
Priority	Low-Level	
Primary Actor	Patient	
Supporting Actors	None	
Pre-Conditions	User needs to be registered in the application	
Trigger		
Main flow	**Actors**	**System**
	1) User selects “View Doctors List” button	2) Application displays the list of doctors registered in the system.

	3) User finds a list of doctors to contact.	4) Contact details of the selected doctor are displayed.
Exception flow	**Actors**	**System**
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	Doctor is contacted informing the user's conditions.	

Table 4. 6: Use Case Table 4

Use Case Name	View Patients List	
Use Case ID	UC-005	
Description	List of patients identified with a heart disease risk.	
Priority	Medium-Level	
Primary Actor	Doctor	
Supporting Actors	None	
Pre-Conditions	User needs to be registered in the application	
Trigger	The user launches the application	
Main flow	**Actors**	**System**
	1) Doctor Selects the "View Patients List" Option 3) Doctor checks the patient's report thoroughly. 4) Doctor recommends the user necessary instructions.	2) Application displays the list of patients grouped as risky. 5) Application instructs the user to follow the necessary steps.
Exception flow	**Actors**	**System**
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	Patient is given a set of instructions to follow.	

Table 4. 7: Use Case Table 5

Use Case Name	View Prediction Result	
Use Case ID	UC-006	
Description	The prediction result of the patient	
Priority	High-Level	
Primary Actor	Patient	
Supporting Actors	None	
Pre-Conditions	User needs to have filled out the heart disease prediction page query	
Trigger	Displays the prediction value	
Main flow	**Actors**	**System**
	1) User fills the parameters from and proceeds.	2) Prediction output is displayed. 3) Recommends a doctor if the prediction outcome is negative.
Exception flow	**Actors**	**System**
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	Recommends a doctor based on their result.	

Table 4. 8: Use Case Table 6

Use Case Name	Doctor Recommendations	
Use Case ID	UC-007	
Description	Recommends doctors for patients after the prediction	
Priority	Medium-Level	
Primary Actor	Patient	
Supporting Actors	Doctor	
Pre-Conditions	User needs to have filled out the heart disease prediction page query	
Trigger	List of Recommended doctors are displayed	
Main flow	**Actors**	**System**
	1) Patient Selects the "View Doctors List" Option 3) Users have the option to select a doctor they prefer.	2) Application displays the list of doctors registered in the system. 4) Upon user's preference, the application hands out the doctor's contact details.
Exception flow	**Actors**	**System**
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	User contacts the doctor.	

Table 4. 9: Use Case Table 7

Use Case Name	Maintaining The System	
Use Case ID	UC-008	
Description	Developers are identifying and fixing the issues.	
Priority	High-Level	
Primary Actor	Development Team	
Supporting Actors	None	
Pre-Conditions		
Trigger		
Main flow	**Actors**	**System**
Exception flow	**Actors**	**System**
Alternate flow	**Actors**	**System**
Exclusions	None	
Post Conditions	Application issues are addressed.	

Table 4. 10: Use Case Table 8

4.7 Functional Requirements (with prioritization)

- Critical – The requirements that are critically needed in the successful completion
- Desirable – The requirements that can add value, but are not required immediately
- Luxury – The requirements that would add luxury to the system

Requirements list		**Priority Level**	**Description**
FR1	Sign up and Login	Critical	User must login to the system by entering username and password. If the user is a new user, should create an account by entering personal details, username and password.
FR2	Asking questions regarding patient heart disease	Critical	After entering details, the system will ask a few questions regarding his heart disease.
FR3	Predict the heart disease	Critical	The system will predict whether the patient suffering from a heart disease or not
FR4	Giving feedback to the user	Critical	The user can see the feedback of their diseases.
FR5	Recommending a cardiologist	Desirable	The system will recommend a cardiologist according to their risk level and type of the heart disease.
FR6	Doctor can insert patient's data and updating the system	Desirable	The system allows doctors to enter information about a patient such as the patient's name, illness, medication provided, arrival date, and cost.

FR7	Doctors can view the data when they needed	Desirable	This system allows the doctors to view previous data when they needed.
FR8	Accessing the system by doctors	Desirable	The doctor can access the system by entering a username and ID.
FR9	Predict the heart disease using ECG	Desirable	The system should be able to predict the heart disease using ECG
FR10	Patient can contact the doctors via online platform	Luxury	When the system recommending a doctor, patient can contact the doctor via online platform
FR11	Searching for a patient or a doctor by their name or user id.	Luxury	System allows users to search a patient or a doctor by their name or user id.

Table 4. 11: Functional requirements

4.8 Non-Functional Requirements

Requirement		Description
NF1	Security	The system will understand the user by his username and ID.
NF2	Maintainability	System will track all errors and keep a log of them, and technology provides efficient data backup.
NF3	Performance	When patient information is validated, the system accepts it in less than one second.
NF4	Capacity	The system must be able to handle at least 1000 users at the same time.
NF5	Reliability	The system will be always available.

Table 4. 12:Non-Functional requirements

4.9. Chapter Summary

This chapter has explained about the System Requirement Specification. This section covered stakeholder analysis, survey questions with discussion, user case diagrams, as well as functional and non-functional requirements.

Chapter 5: Social, Legal, Ethical and Professional Issues

5.1 Chapter Overview

When working on a project, almost always have to deal with a variety of challenges and issues. This chapter will discuss some social issues, legal issues, ethical issues and professional issues related to our project. On the other hand, the unique aspect is that in order to construct a successful project, firstly one needs to identify those issues and then decrease or eliminate them. So here, it will mention how to mitigate those issues in order to build a successful project.

5.2 SLEP Issues and Mitigation

5.2.1. Social Issues

At first, the application (UrHeart) will only be available in the English language. As a result of this, users who do not know the English language cannot use this application for prediction purposes and will have to wait until a later version adds multilingual support. According to BCS, access to communication and information should be a basic human right for all people including the elderly and disabled people also. But there are several social issues identified related to our proposed application. The users who do not know much technology and the idea about using this kind of system or mobile applications, also cannot use this system. Disable people such as blind people are also unable to use this kind of application for their needs.

5.2.2 Legal Issues

There are several legal issues faced when collecting the information for this project and finding the dataset. Some websites didn't allow to access the information without paying. Sometimes when searching for information from several websites, must need things like a license to log in. On the other hand, unauthorized access to those websites can lead to legal issues. In addition to that, if the datasets included some incorrect information, it may result in legal issues due to the inconvenience caused to app users. The information gathered from the questionnaires was all done with the user's privacy in mind, and no personal information was collected in the process. None of the above-mentioned information was used for any illegal purposes.

5.2.3 Ethical Issues

There are several ethical issues that occurred when collecting the information for this project and finding the dataset. The issues which are related to privacy and security are very crucial, especially our application which consists of sensitive information like patient data related to diseases. Because sometimes information and dataset consist of some incorrect information. So it will mainly affect the accuracy of the application. Questionnaires were well informed about the nature of the project and the usage of their contribution to the project to make an intelligent, proper decision. Only a secured private version control environment and a private Google Drive account were used throughout the whole project. None of the dissertation or the results is plagiarized. When taking information from the available sources, it is also properly cited.

5.2.4 Professional Issues

The data sets, the questionnaires, and the software that used will be professionally implemented. Before starting into the app every user must answer some primary questions that are basically needed for the program guidelines will be provided through the app. Questionnaires' responses were collected anonymously but the primary questions won't be anonymous but that data will be only available for the respective doctor and the user. Lack of contribution of the group members was one of the critical issues which lead to team demotivation. As a result, project targets and deadlines could be unachievable. To overcome this issue, team members should be responsible for each task assigned to them.

5.4 Chapter Summary

This chapter provided some issues which are related to the proposed project such as social issues, legal issues, ethical issues, and professional issues. In addition to that, discussed about how to mitigate those issues in order to build a successful project.

Chapter 6: System Architecture & Design

6.1 Chapter Overview

The previous chapter explained the social, legal, ethical, and professional issues. This is the final chapter, and it will discuss the System Architecture and Design for our project. The class diagram, sequence diagram, suggested UI design for our project, and system process flow diagram also will be discussed in this chapter.

6.2 System Architecture Design

The below figure shows the system architecture design for the proposed project (UrHeart heart disease prediction system). It mainly consists of three layers which are the data collection layer, interaction UI layer and data processing layer. Among all these layers, the real-time data collection layer contains some vital information about patients and doctors and ECG image data. The interactive UI layer includes gathered data from the application and displaying results from the system. The data processing layer is the layer that handles the data which is collected from the application and pre-processing that gathered information. The data preprocessing layer interacts with other layers to handle missing data. After data pre-processing, those data are used for training and testing purposes. Machine learning algorithms like the K-nearest neighbors' algorithm are used to predict heart disease. After predicting the heart disease, that data processing layer interacts with the UI layer in order to display predicted results.

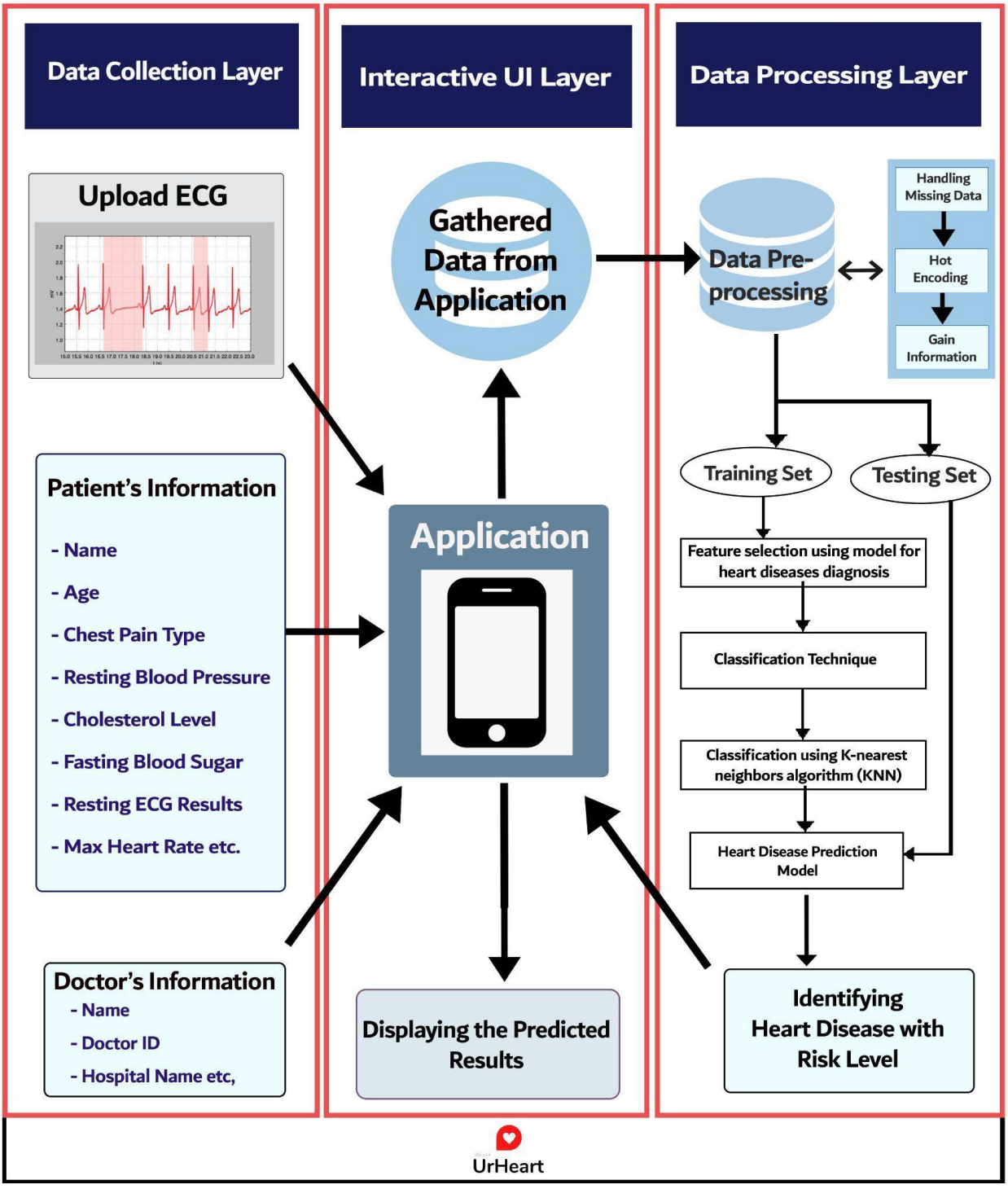


Figure 6. 1: System Architecture Design

6.3 System Design

6.3.1 Class Diagram

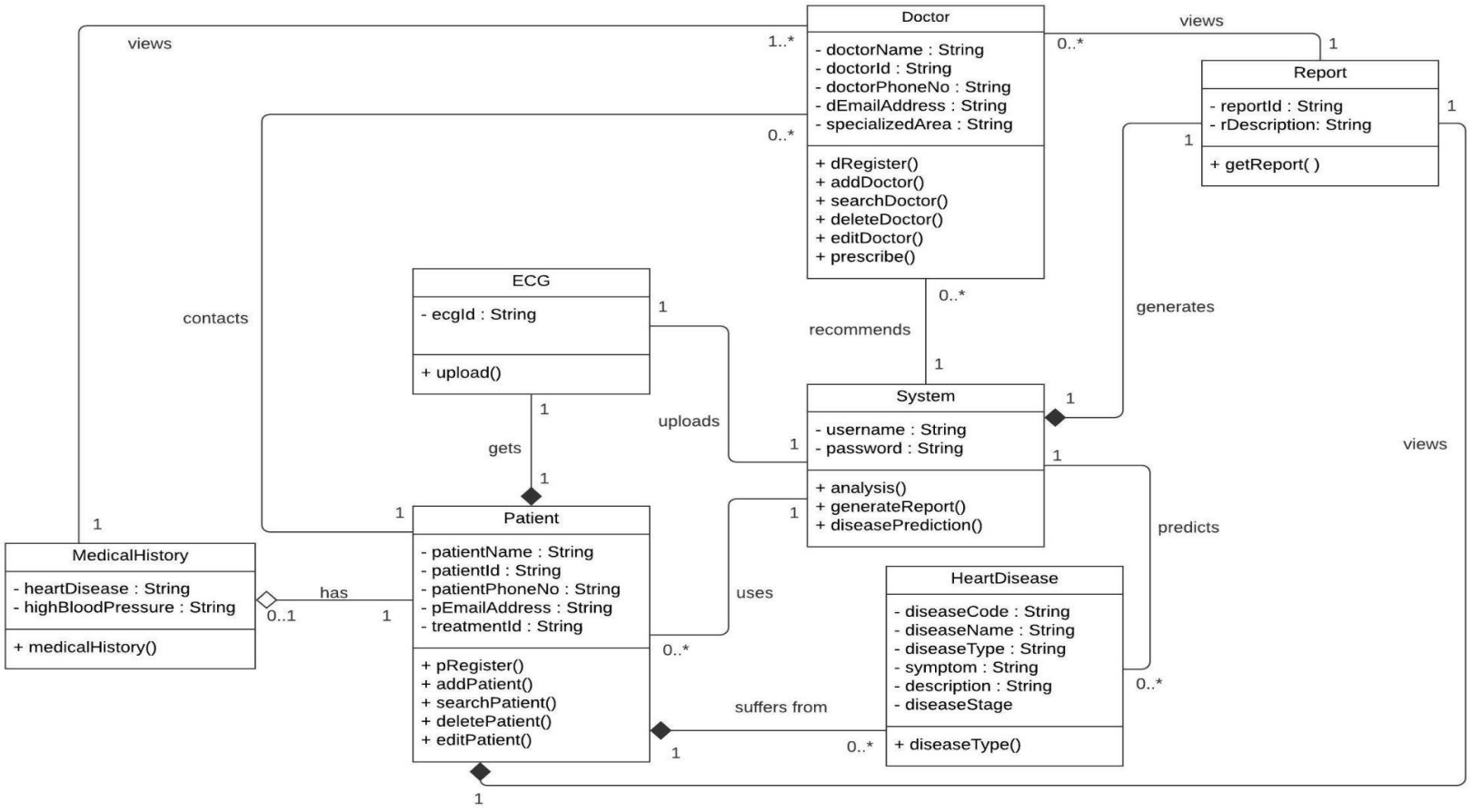
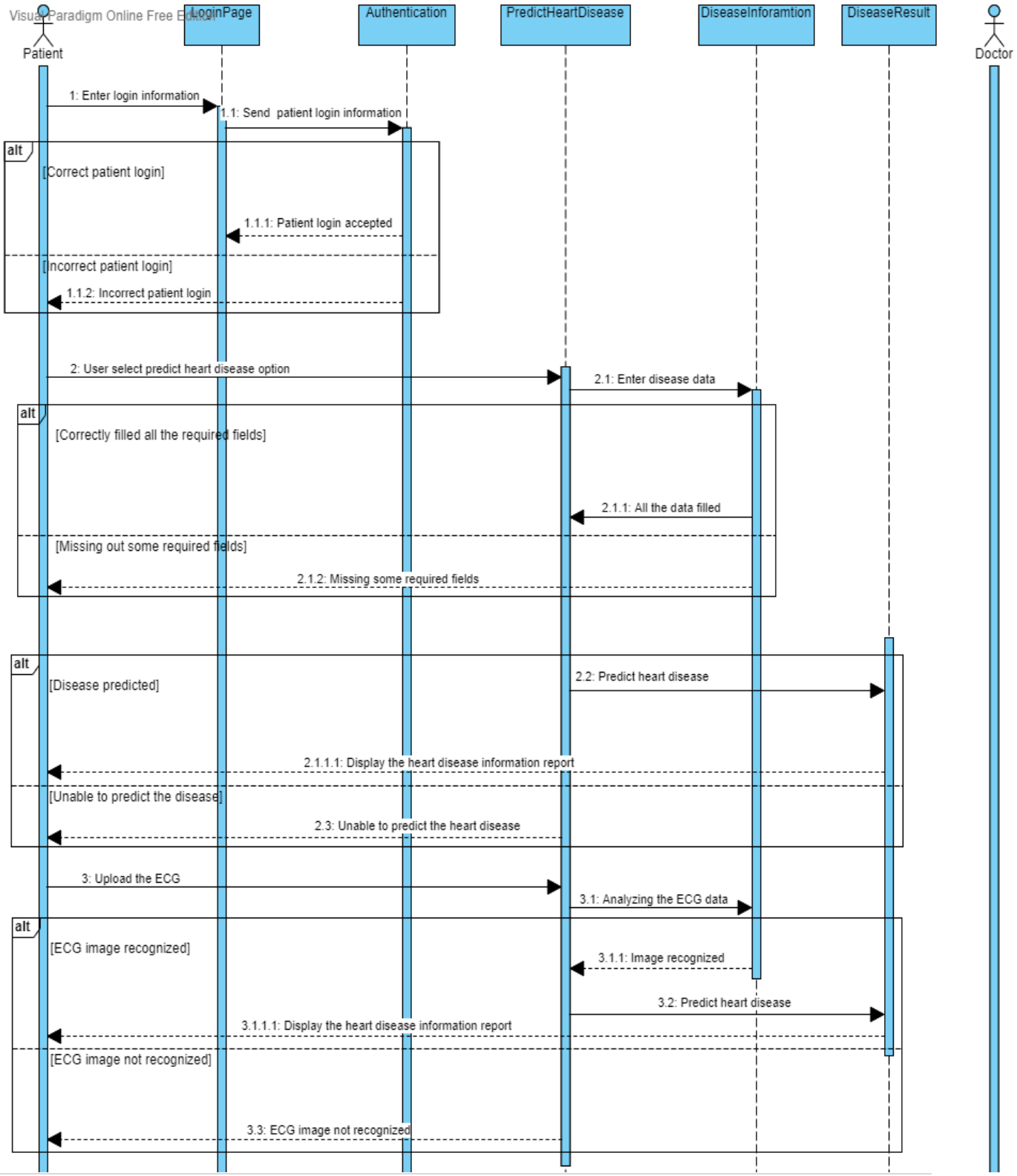


Figure 6. 2: Class Diagram

6.3.2 Sequence Diagram



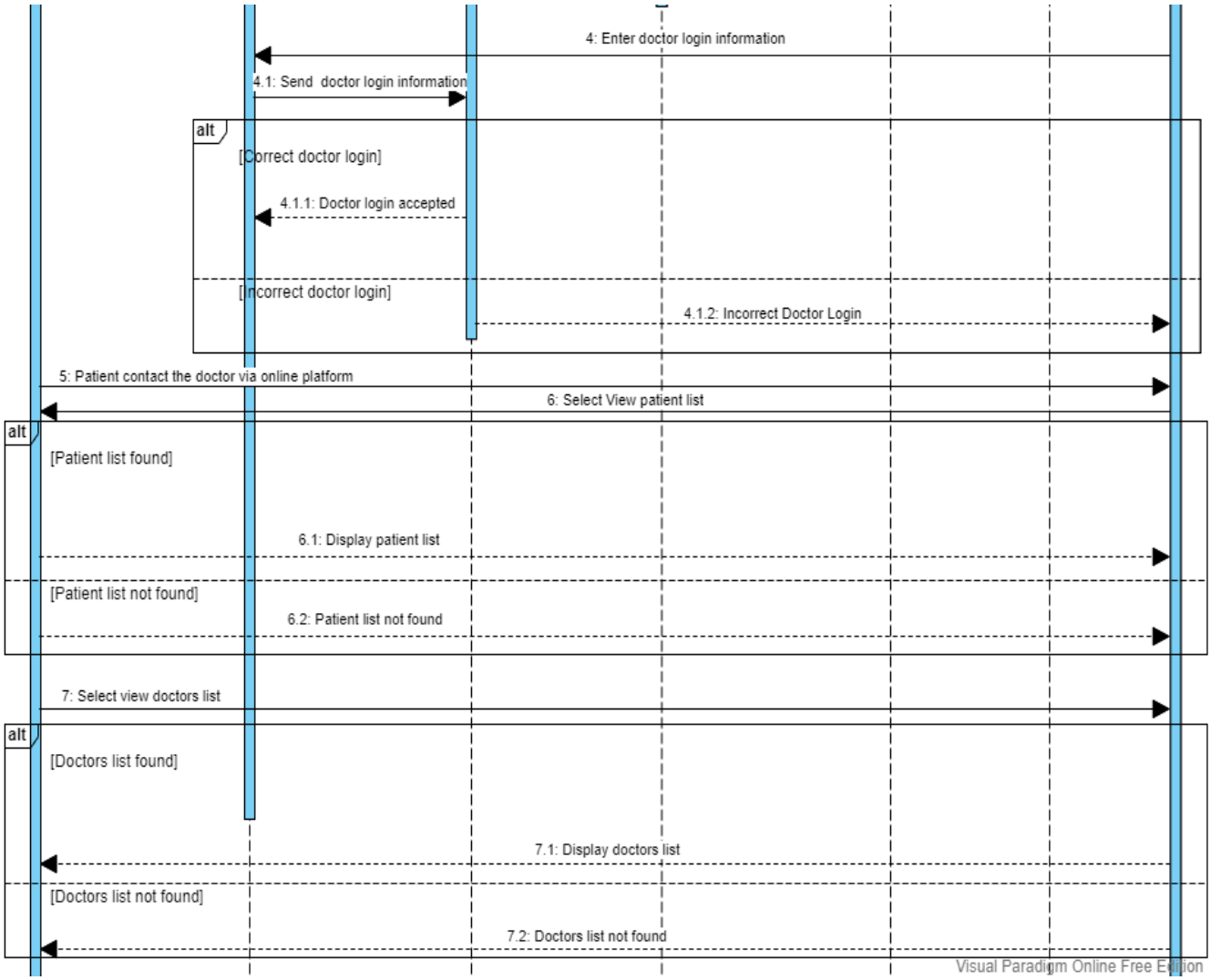


Figure 6. 3: Sequence Diagram

6.3.3 UI Design

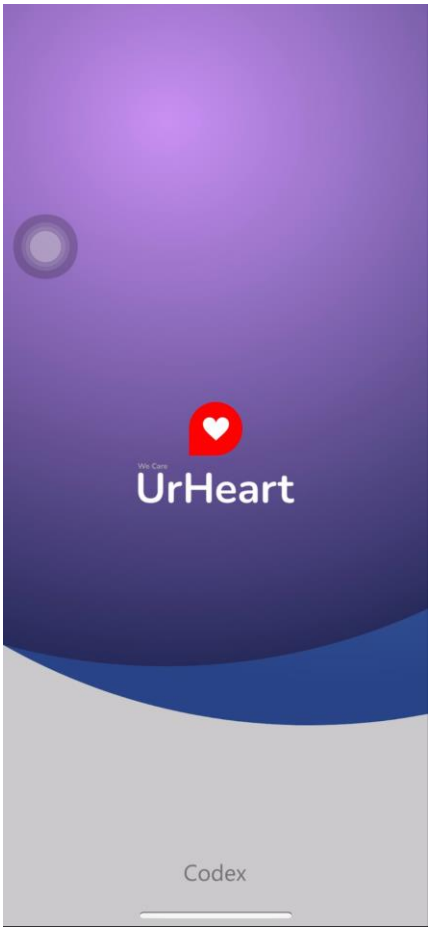


Figure 6. 4: Loading Page

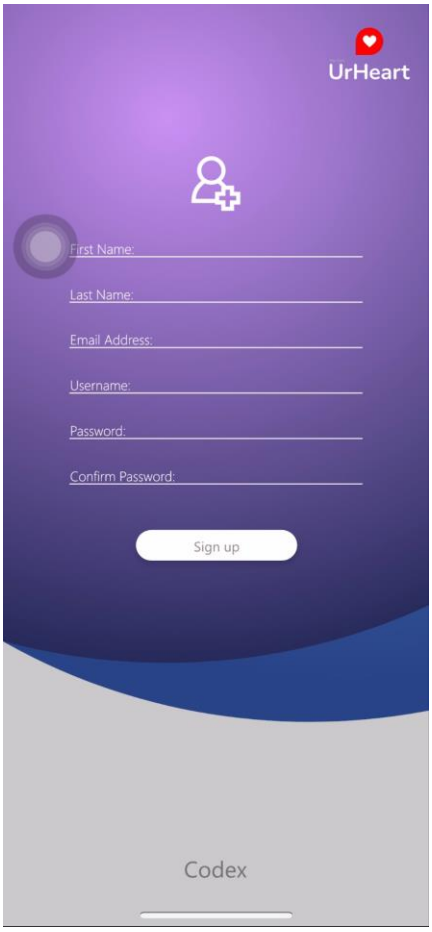


Figure 6. 5: Signup Page

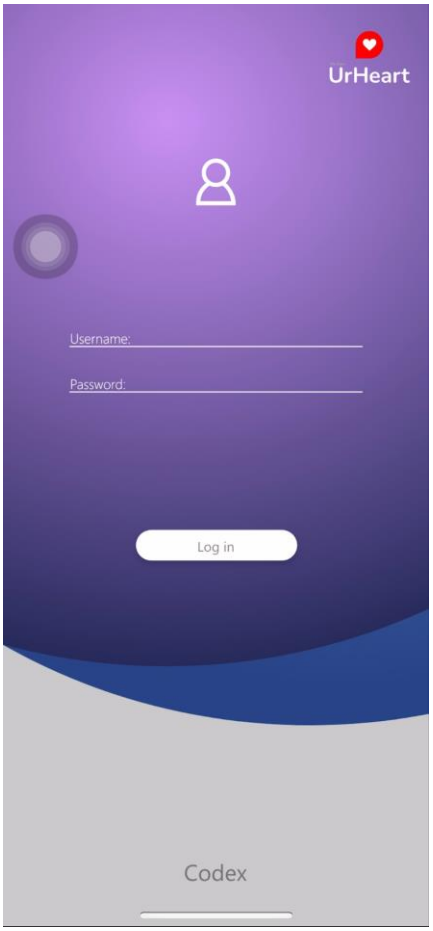


Figure 6. 6: Login Page

Figure 6. 7: Home Page

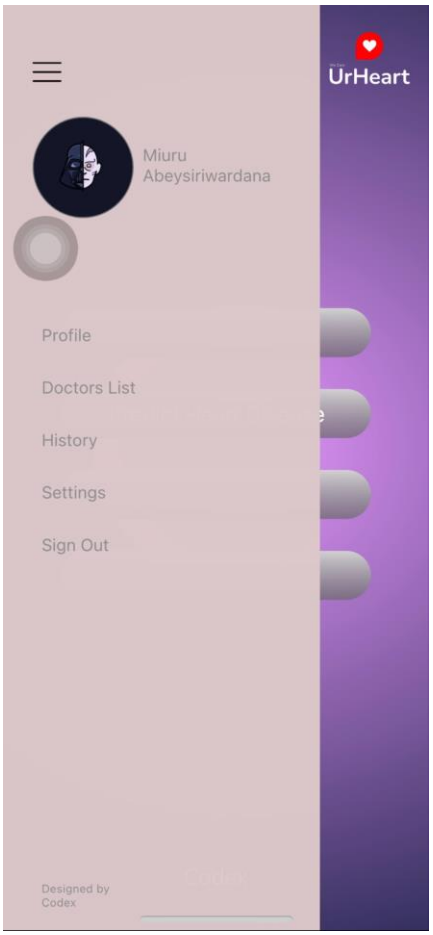


Figure 6. 8: Side Bar

Figure 6. 9: User Profile

Figure 6. 10: Doctors List

Figure 6. 11: Prediction Data

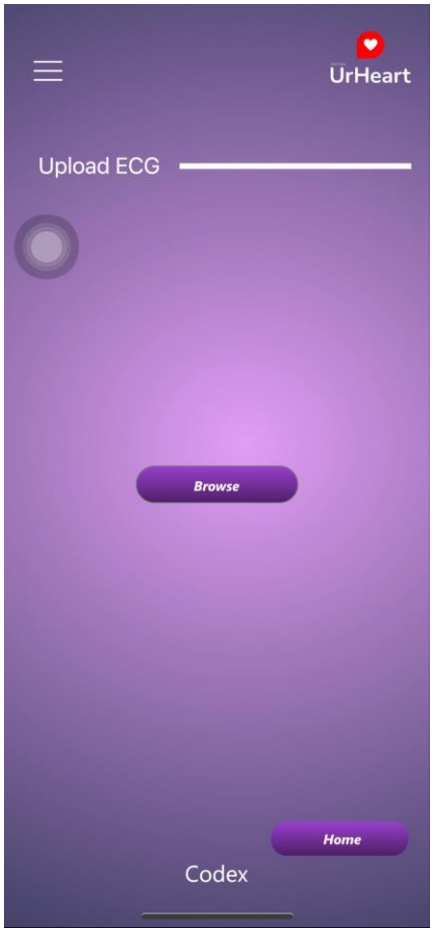


Figure 6. 12: Upload ECG page

6.3.4 Process flow chart

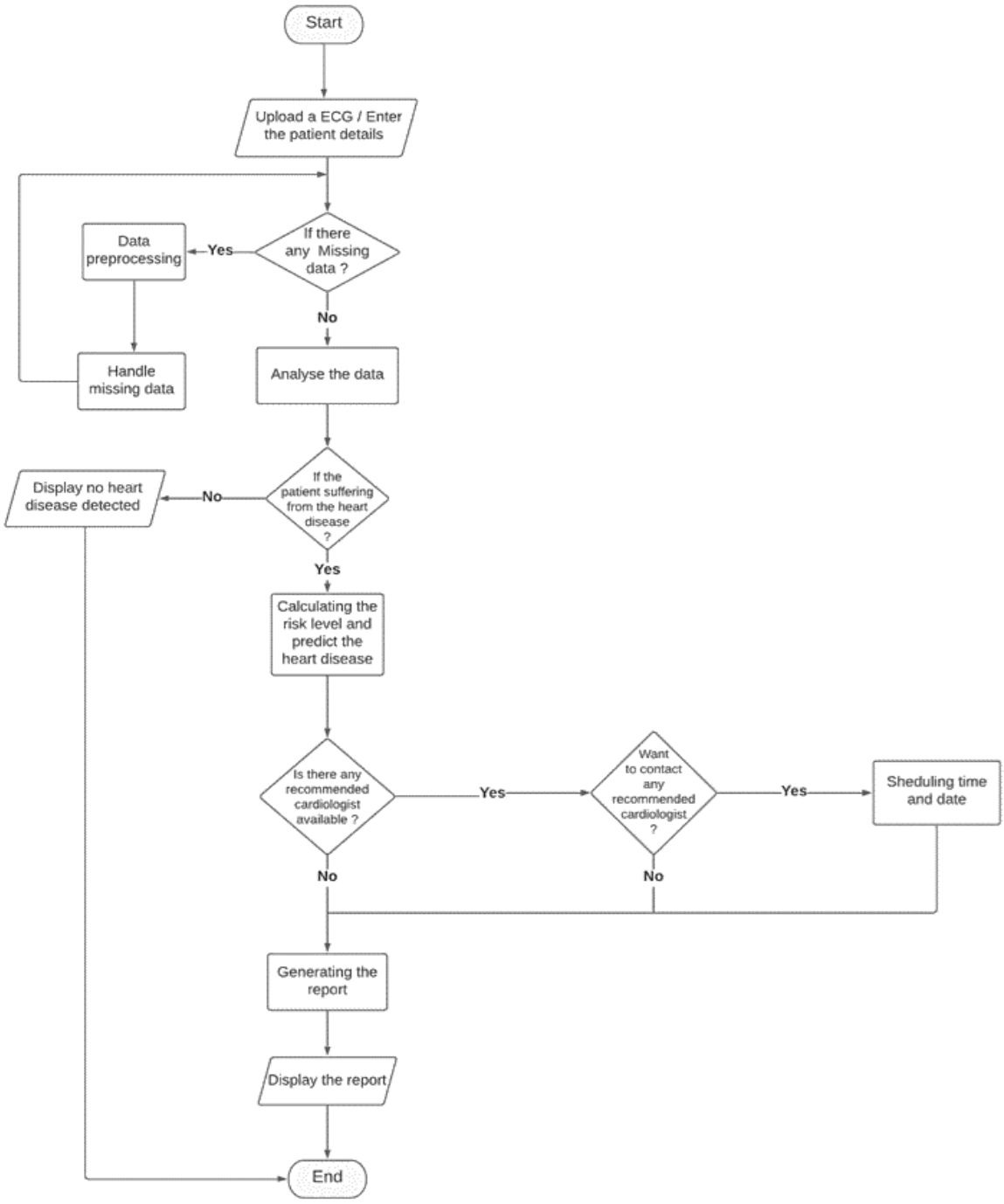


Figure 6. 13: Process Flow Chart

6.4 Chapter Summary

This final chapter provided an overall System Architecture and Design for the proposed project. The class diagram, sequence diagram, suggested UI design, and system process flow diagram has been discussed in this chapter.

References

Abood, R.A. and Alalwany, A.A.Z. (2021). Investigating the performance of Iraqi EFL researchers in writing quantitative and qualitative research. *Journal of Language and Linguistic Studies*, 17 (4). Available from http://jlls.org/index.php/jlls/article/view/3474 [Accessed 8 December 2021].

Arul Jothi, K., Subburam, S., Umadevi, V. and Hemavathy, K., 2021. Heart disease prediction system using machine learning. Materials Today: Proceedings,.

A.T.Matthias, D.K.N.deSilva, J.Indrakumar, S.B.Gunatilake, (2018) Physical activity levels of patients prior to acute coronary syndrome – Experience at a tertiary care hospital in Sri Lanka | Elsevier Enhanced Reader. Available from https://doi.org/10.1016/j.ihj.2017.08.020 [Accessed 30 December 2021].

Bavani, B., Rajini, S., Josephine, M. and Prasannakumari, V., 2019. Heart Disease Prediction System based on Decision Tree Classifier. Journal of Advanced Research in Dynamical and Control Systems, 11(10-SPECIAL ISSUE), pp.1232-1237.

B.Bontchev, E.Milanova (2020) On the Usability of Object-Oriented Design Pattern [Online]Available at: https://cit.iict.bas.bg/CIT-2020/v-20-4/10341-Volume20_Issue_4-03_paper.pdf [Accessed 23 December 2021].

Cardiovascular diseases. (no date).[Web Page] Available from https://www.who.int/westernpacific/health-topics/cardiovascular-diseases [Accessed 30 December 2021].

Chandra, V., 2015. Comparison between Various Software Development Methodologies. International Journal of Computer Applications, 131(9), pp.7-10.

Chen, S., Webb, G.I., Liu, L., Ma, X., 2020. A novel selective naïve Bayes algorithm. Knowledge-Based Systems 192, 105361. [Online]Available at: https://doi.org/10.1016/j.knosys.2019.105361 [Accessed: 25 November 2021]

Corus, D., Dang, D.-C., Eremeev, A.V., Lehre, P.K., 2018. Level-Based Analysis of Genetic Algorithms and Other Search Processes. IEEE Transactions on Evolutionary Computation 22, 707–719. [Online]Available at: https://doi.org/10.1109/TEVC.2017.2753538 [Accessed: 23 November 2021].

Eshlaghy, A. et al. 2011., A Classification of Qualitative Research Methods. [Online]Available at: A-Classification-of-Qualitative-Research-Methods.pdf (researchgate.net) [Accessed: 8 December 2021].

Fatima, M. and Pasha, M., 2017. Survey of Machine Learning Algorithms for Disease Diagnostic. Journal of Intelligent Learning Systems and Applications, 09(01), pp.1-16.

Janosi, Andras, Steinbrunn, William, Pfisterer, Matthias, Detrano, Robert & M.D., M.D.. (1988). Heart Disease. UCI Machine Learning Repository.

J. Murtagh and E. Sobiesk, "Exploring the ethical, legal, and social aspects of the computing discipline," Proceedings Frontiers in Education 1997 27th Annual Conference. Teaching and Learning in an Era of Change, 1997, pp. 474-477 vol.1, doi: 10.1109/FIE.1997.644927 https://ieeexplore.ieee.org/abstract/document/644927

Kaminsky, L. and Tuttle, M., 2015. Functional Assessment of Heart Failure Patients. Heart Failure Clinics, 11(1), pp.29-36.

Khan, Ali Haider; Hussain, Muzammil (2021), “ECG Images dataset of Cardiac Patients ”, Mendeley Data, V2, doi: 10.17632/gwbz3fsgp8.23.2011 Computing in cardiology A.H.Chen, S.Y.Huang, P.S.Hong, C.H.Cheng , E.J.lin

Khan, D., 2021. Cardiovascular diseases. [online] Who.int. Available at: https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1 [Accessed 4 November 2021].

Lee, D., 2021. Women and heart disease: Frequently asked questions. [online] University of Iowa Hospitals & Clinics. Available at: https://uihc.org/health-topics/women-and-heart-disease-frequently-asked-questions [Accessed 1 November 2021].

Machine Learning Decision Tree Classification Algorithm - Javatpoint (no date) *www.javatpoint.com*. [Online]Available at: https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm [Accessed: 21 November 2021].

Marimuthu, M., Abinaya, M., S., K., Madhankumar, K. and Pavithra, V., 2018. A Review on Heart Disease Prediction using Machine Learning and Data Analytics Approach. International Journal of Computer Applications, 181(18), pp.20-25.

Matthias, A., de Silva, D., Indrakumar, J. and Gunatilake, S., 2018. Physical activity levels of patients prior to acute coronary syndrome – Experience at a tertiary care hospital in Sri Lanka. Indian Heart Journal, 70(3), pp.350-352.

McFarland, J., 2017. Heart Attack and Stroke Symptoms. [Online]Available at: https://www.heart.org/en/health-topics/heart-failure/what-is-heart-failure [Accessed 4 November 2021].

Mihaljevic, T., 2021. *Heart Failure.* [Online]Available at: https://my.clevelandclinic.org/health/diseases/17069-heart-failure-understanding-heart-failure [Accessed 29 October 2021].

Muhammad, Y., Tahir, M., Hayat, M. et al. Early and accurate detection and diagnosis of heart disease using an intelligent computational model. Sci Rep 10, 19747 (2020). https://doi.org/10.1038/s41598-020-76635-9

number8. 2022. Kanban Versus Scrum in Agile Software Development. [online] Available at: https://number8.com/kanban-versus-scrum [Accessed 17 January 2022].

Peterson, L.E., 2009. K-nearest neighbor. Scholarpedia 4, 1883. [Online]Available at: https://doi.org/10.4249/scholarpedia.1883. [Accessed 17 November 2021].

Rajendran, N. and Vincent, D., 2021. Heart Disease Prediction System using Ensemble of Machine Learning Algorithms. Recent Patents on Engineering, 15(2), pp.130-139.

Rose, G.B. (1991). SSADM-the open methodology. *1991 IEE Colloquium on an Introduction to Software Design Methodologies (Ref. No: 1991/181)*. December 1991. 6/1-6/5. [Online]Available at: https://ieeexplore.ieee.org/abstract/document/182227 [Accessed 24 December 2021].

Saeedi, K. and Visvizi, A., 2021. Software Development Methodologies, HEIs, and the Digital Economy. Education Sciences, 11(2), p.73.

Saxenab, P. and Sharma, R., 2016. Efficient Heart Disease Prediction System. [online] Uttar Pradesh, Noida, India: Procedia, pp.962, 963, 964. Available at: <http://surl.li/anfvj> [Accessed 28 October 2021].

Singh, P., Singh, S. and Pandi-Jain, G., 2018. Effective heart disease prediction system using data mining techniques. International Journal of Nanomedicine, Volume 13, pp.121-124.

Smith, B. (2011). Object-Oriented Programming. In: Smith, B. (ed.). *AdvancED ActionScript 3.0: Design Patterns*. Berkeley, CA: Apress, 1–25. Available from https://doi.org/10.1007/978-1-4302-3615-3_1 [Accessed 23 December 2021].

TatvaSoft Blog. 2022. Top 12 Software Development Methodologies & its Advantages & Disadvantages. [online] Available at: <https://www.tatvasoft.com/blog/top-12-software-development-methodologies-and-its-advantages-disadvantages/> [Accessed 25 November 2021].

Tripathi, S. F. a. M., 2021. A Low-Sodium Diet Boosts Ang (1–7) Production and NO-cGMP Bioavailability to Reduce Edema and Enhance Survival in Experimental Heart Failure. *Heart Failure: From Molecular Basis to Therapy,* pp. 1-3.

Appendix

Work Breakdown Chart

Task	Codex Group Members				
	Miuru	Ojitha	Tharushi	Madushani	Thariq
Chapter 1 - Introduction					
Chapter Overview				✔	
Problem Background				✔	
Problem Statement				✔	
Research Gap	✔				
Research Questions					✔
Research Aim				✔	
Project Scope		✔			
Requirements	✔		✔		✔
Objectives	✔				
Resource Requirements	✔	✔			✔
Chapter Summary					✔
Chapter 2 - Literature Review					

Chapter Introduction	✔				
Existing Work		✔	✔	✔	
Proposed Solution Overview					✔
Tools and techniques	✔				✔
Chapter Summary				✔	
Chapter 3 - Methodology					
Chapter Overview				✔	
Research Methodology					✔
Development Methodology				✔	
Design methodology					✔
Evaluation methodology					✔
Project Management Methodology				✔	
Team Work Breakdown Structure (WBS)				✔	
Gantt chart diagram				✔	
Usage of Project Management and Collaboration Software In the project				✔	✔
Risks and Mitigation				✔	

Chapter Summary				✔	
Chapter 4: System Requirements Specification (SRS)					
Chapter Overview	✔				
Onion Model	✔				
Stakeholder Descriptions	✔				
Selection of Requirement Elicitation Techniques/Methods	✔				
Discussion/ Analysis of Results	✔				
Use Case Diagrams	✔				
Use Case Descriptions	✔				
Functional Requirements				✔	
Non-Functional Requirements			✔		
Chapter Summary				✔	
Chapter 5: Social, Legal, Ethical and Professional Issues					
Chapter Overview			✔	✔	
SLEP Issues and Mitigation		✔	✔	✔	
Chapter Summary		✔			

Chapter 6: System Architecture & Design					
Chapter Overview				✔	
System Architecture Design				✔	
Class Diagram				✔	
Sequence Diagram				✔	
UI Design	✔				
Process flow chart				✔	
Chapter Summary				✔	
Logo Design					✔