A Major Project Final Report on

E-Health Care

Submitted in Partial Fulfillment of the Requirements for

the degree of **Bachelor in Information Technology**

Under Pokhara university

Submitted by:

Pawan Upreti, 161517

Sushil Phuyal, 161545

Pradip Thapa, 161547

Pramod Lamichhane, 161551

Under the supervision of

Mr. Anish Chapagain

Date: 5, July, 2021

Nepal College of Information Technology

Affiliated to Pokhara University

ACKNOWLEDGEMENT

After a rigorous and continuous quantum of united efforts we are glad to present this report as an outcome of the major project report on “E Health Care”. We express our sincere gratitude to those helping hands that come forward and directed us towards the success of the whole project with the desired output.

We extend our genuine thankfulness to our project supervisor Mr. Anish Chapagain for his valuable and great support and guidance in all the happenings regarding to the project. We must express out thanks to Dr. Roshan Chitrakar, project coordinator for his encouragement and guidance throughout the project.

We are indebted to the department of information technology for providing support to add on out venture. Last, but not the least, we would like to thank our teachers and colleagues who have been knowingly or unknowingly the part of this project and lent support and views during the entire development time.

ABSTRACT:

Health care services (web based) are demanding tools that are very delicate and face vast challenges in forms of concepts, tools and implementation that will increase in the years to come. This is due to demographic changes, such as an ageing population and a lack of labor plus ever-growing technical tools. For this E-Health care is viewed as an important way to meet these challenges. The term 'E-Health Care' can be defined as 'The application of information, communication, computing, and sensing technologies across the entire range of functions and processes constituting the practice and delivery of health care services'. E-Health care is enabled through integrated applications in the healthcare environment, and includes technologies related to computing, communication and sensing.

E-health care information is widely used with different meanings and purposes. In our work, we have developed E-health care application specially targeted for Heart diseases, Diabetes, Pneumonia patient. We have developed a classification model with the help of machine learning algorithms like random forest, KNN algorithm and CNN algorithm which will record the symptoms felt by person and according to that symptoms it predicts the possibility of the diseases that the person is suffering. This application is built upon popular framework Django from Python Programming Language. The database that will be used throughout the web application is MYSQL. It saves time as well as makes it easy to get a warning about your health before it's too late. E-health care model allows users to communicate with health care professionals, to access medical records, to research health information, and to engage in person-to person exchange of text. This will accelerate the much-needed transformation of our health care systems and sustain access, affordability and quality for all in the near future.

Keywords: Python, Django, KNN Algorithm, Random Forest & CNN Algorithm, MYSQL

TABLE OF CONTENTS

1. INTRODUCTION 1
1.1 PROBLEM STATEMENT: 1
1.2 PROJECT OBJECTIVE 1
1.3 SIGNIFICANCE OF THE STUDY 1
1.4 SCOPE 2
1.5 LIMITATION 2
2. LITERATURE REVIEW 3
3.REQUIREMENT ANALYSIS 6
3.1 SOFTWARE SPECIFICATION 6
3.1.1 PYTHON 6
3.1.2 SKLEARN 6
3.1.3 PANDAS 6
3.1.4 NUMPY 6
3.1.5 SEABORN AND MATPLOTLIB 6
3.1.6 DJANGO 6
3.1.7 MYSQL 6
3.1.8 HTML 7
3.1.9 CASCADING STYLE SHEET(CSS) 7
3.1.10 JAVASCRIPT 7
3.1.11 BOOTSTRAP 7
3.2 FUNCTIONAL REQUIREMENT 7
3.3 NON-FUNCTIONAL REQUIREMENT 8
4. PROJECT DETAILS AND UML DIAGRAMS 9
4.1 USE CASE DIAGRAM 9
4.2 CONTEXT DIAGRAM 11
4.3 CLASS DIAGRAM 11
4.4 ENTITY RELATIONSHIP DIAGRAM 13
4.5 DATA FLOW DIAGRAM 15
5. PROPOSED METHODOLOGY 17
5.1 SOFTWARE DEVELOPMENT LIFECYCLE 17
5.1.1 REQUIREMENT ANALYSIS 18
5.1.2 DESIGN PHASE 18
5.1.3 CODING PHASE 18

5.1.4 TESTING PHASE 18
5.2 EXPLORATORY DATA ANALYSIS 18
5.3 ALGORITHM USED 22
5.3.1 K-NEAREST NEIGHBOUR (KNN) ALGORITHM 22
5.3.2 VISUALIZATION OF TRAINING IMAGES 29
5.3.3 RANDOM FOREST ALGORITHM 33
5.4 SECURING DATA IN WEB 35
6. TESTING 36
6.1 TESTING PLAN 36
6.2 TESTING METHODS 36
6.2.1 BLACK BOX AND WHITE BOX TESTING 36
6.2.2 UNIT TESTING 37
6.2.3 INTEGRATION TESTING 38
6.3 TEST CASES 38
7. PROJECT TASK AND TIME SCHEDULE 41
7.1 GANTT CHART 42
8. CONCLUSION 43
9. FURTHER WORKS/FUTURE RECOMMENDATION 43
10. BIBLOGRAPHY 44

LIST OF FIGURES

Figure 1 : Use case diagram 10
Figure 2 : Context Diagram 11
Figure 3 : Class Diagram 12
Figure 4 : ER Diagram 14
Figure 5 : DFD for heart problem prediction 15
Figure 6 : DFD for diabetes problem prediction 15
Figure 7 : DFD for disease prediction 16
Figure 8 : DFD for pneumonia prediction 16
Figure 9 : Iterative model of software development life cycle 17
Figure 10 : Information about heart.csv 19
Figure 11: Value count for normal condition and heart problem 19
Figure 12: Heart problem analysis with respect to independent variable 20
Figure 13: Heart disease in function of age and max heart rate 21
Figure 14 : Data Processing of heart.csv 21
Figure 15 : New data point and category 22
Figure 16 : Calculation of Euclidean distance between two points 23
Figure 17 : Selection of category for new data point 23
Figure 18 : K-NN accuracy for heart problem prediction 25
Figure 19 : Information about diabetes.csv 25
Figure 20 : Value count for normal condition and diabetes problem 26
Figure 21 : Diabetes disease in function of age and maximum diabetes rate 26
Figure 22 : Total count of individual data 27
Figure 23 : K-NN accuracy for diabetes disease prediction 29
Figure 24 : Representation of chest x-ray showing pneumonia and normal report 29
Figure 25 : Working of CNN model 30
Figure 26 : Accuracy of CNN algorithm 31
Figure 27 : Analysis of disease datasets 32
Figure 28 : Random Forest Algorithm 33
Figure 29 : Accuracy of RF algorithm 34
Figure 30 : Gantt chart 42

LIST OF TABLES

Table 1 : Test Cases 40
Table 2: Project task and time schedule.....41

LIST OF ABBREVIATIONS

DFD	Data Flow Diagram
ER	Entity Relationship
HTML	Hypertext Markup Language
SQL	Structured Query Language
MD5	Message Digest 5
CNN	Convolution Neural Network
ANN	Artificial Neural Network
UML	Unified Modeling Language
K-NN	K-Nearest Neighbor

1. INTRODUCTION

The E Health Care system is an end user support and online consultation project. It might have happened so many times that you or someone yours need doctors help immediately, but they are not available due to some reason. E Health Care is a web-based application that allows patient to get instant guidance on their health issues through an intelligent health care application online. This application is fed with various features that allows patient to choose symptoms, predict the disease based on it and suggest the doctors for consultation. It also allows patient to upload information of heart report, diabetes report and a chest x-ray so that the system will able to predict heart issue, diabetes and pneumonia. This system also aims to connect patients to available doctors for consultation and also allows doctors to email subscriptions. This system aims to build an environment that will prove helpful to urgent cases that do not reach hospital, for emergency cases that do not have doctors in area, during late night emergencies and also for preliminary examination of patients and allow a doctor to schedule an appointment. In this section, problem and motivation, objectives, project scope and limitation will be described in detail.

1.1 PROBLEM STATEMENT:

- Patient has to suffer a lot going from one doctor to another to identify their problem.
- Wastage of money as well as time of the patient in the intermediate state before identifying their problem.

1.2 PROJECT OBJECTIVE

The main objective for this project on E-Health care is as follows:

- Recommend the specialized doctor according to the disease predicted by the system as per patient's statements.
- To provide search facility according to diseases and their corresponding specialty.
- Take appointment with specific doctor as per patient need.

1.3 SIGNIFICANCE OF THE STUDY

- E-Health Care medical records with specialized software programs that can increase the quality of patient care, reduce unnecessary medical tests, and directly connect with professional doctors.
- E-Health is the cost-effective and protected use of facts and communication technologies in favor of health and health-related fields.
- It provides quick access to patient records and information for efficient health care.

- Reduced paper work, reduced duplication of costs etc. thus reducing the cost of health care.
- Better health care by improving all aspects of patient care, including safety, effectiveness, patient centeredness, communication, education, timeliness, efficiency, and equity.
- E-Health care saves time.

1.4 SCOPE

- Increase employment opportunities for both technical and medical fields.
- Quick access to patient records and information for efficient health care.
- Reduced medical errors and better clinical decision making.
- Patient can search for doctor's help at any point of time.
- Patient can talk checkup for problem and get required medical prescription.
- Doctors can handle emergency situation by providing primary help, till the patient can be taken to the hospital.
- Better health care by improving all aspects of patient care, including safety, effectiveness, patient centeredness, communication, education, timeliness, efficiency, and equity.

1.5 LIMITATION

- Trouble in learning and using the software.
- It might be time-consuming to bring up-to-date the record comprehensively
- The system is not fully automated, it needs doctors for full diagnosis.
- Will provide inaccurate results if the data is fed incorrectly
- The systems being used must be user friendly otherwise they will not be easily accepted and not used to their full capacity.
- Potential privacy and security issues.

2. LITERATURE REVIEW

For nearly 20 years, a major social problem in Nepal is the difficulty and high cost of getting medical treatment, which is also a concern of people's livelihood. Although many new health care policies continue being proposed, the problem has not been solved fundamentally. A survey found that the most prominent difficulty of getting medical treatment for patients is to find suitable doctors. For safety and security, patients tend to pay high fees only to select big hospitals, regardless of their illnesses. But when all patients choose big hospitals, a series of problems emerge, including long queuing time, medical shortage, high fees and bad service attitudes of medical staff, etc. Personalized doctor recommendation can recommend the most suitable doctor according to each patient's interests and needs. On one hand, it will not delay the illness; on the other hand, it can solve the problems caused by too many people choosing limited number of big hospitals. For example, resource in big hospital is often in shortage while grass-roots medical facilities are idle. Compared with the other algorithms, recommendation system has a brighter future in development which makes people feel more interested in it and start trying to find its rational uses in other areas.

Traditional methods of recommendation system include recommendation based on association rules [1], demographic [2], content [3], collaborative filtering [4] and on network [5], etc. The method based on association rules tries to mine the internal linkages between the items selected by the user find the items that are selected together and then use them in next recommendation. The method based on demographic and content finds similar users and projects to complete the recommendation, respectively. The method based on
collaborative filtering considers users' preference and finds the correlation between users and projects, besides the items we mentioned above. The method based on network completes the recommendation by building network models. However, only one single recommendation method has limitations inevitably, which is not suitable for complex medical diagnosis.

Paper [6] gives the survey for Disease prediction in big data healthcare using extended CNN. This concept is applied in the medical field to implements the hospital. It provides the (i) high accuracy, (ii) high performance, (iii) high convergence speed. To select the particular region and then, analyzed the chronic diseases, that holds the structured data (extracted useful features), the unstructured data is use the CNN technique, so automatically selects the features. The novel CNN is proposed the medical data, and disease risk model is combined this data. The characteristic behavior of this system is selecting the data via previous term. This term is previously applied is possible but not satisfied the disease changes, because disease level is not standard, it is changed in every second. To take the selected data from large number of data

and improves the accuracy by using risk classification term. The proposed system aim is to predict the risk in liver-oriented disease. So, the hospital dataset is related to the liver-oriented disease and it collects only the structured data from liver disease information. In the proposed system is use the disease risk modeling and get the accuracy. But the risk prediction depends on the different feature of medical data with higher accuracy. Its advantages are high accuracy, high performance and high convergence speed and disadvantages are that the risk prediction depends on the different feature of medical data.

The research paper, "Review on Prediction of Diabetes using Data Mining Technique", elaborates about detailed review of existing data mining methods used for prediction of diabetes. It also gives about the types of diabetes disease Type1, type2, and type3.The aim of the diabetes is to predict the diabetes with the help of Data mining methods such as the K-Nearest Neighbor Algorithm, Bayesian Classifier, Naive Bayesian Classifier, Bayesian Network, all the methods are used for prediction of diabetes. This paper also mentions about the effects of diabetes on patients [7]

. The research paper, "A Survey on Naive Bayes Algorithm for Diabetes Data Set Problems", explores about various Data mining algorithm approaches of data mining that have been utilized for diabetic disease prediction. In this paper Classification and Naive Bayes is one of the most used algorithms for the prediction of disease [8]

Rubin et al. [9] developed a CNN model to detect common thorax disease from frontal and lateral chest X-ray images. MIMIC-CXR dataset was used to perform large-scale automated recognition of these images. The dataset was split into training, testing and validation sets as 70%, 20%, 10% respectively. Data augmentation and pixel normalization were used to improve overall performance. Their dual net CNN model achieved an average AUC of 0.72 and o.688 for PA and AP respectively. A deep convolutional neural network to classify pulmonary tuberculosis was developed by Lakhani et al. [10]. Transfer learning models such as Alex Net and Google Net were also used to classify chest X-ray images. The data set was split into training, testing and validation set as 68%, 14.9% and 17.1% respectively. Data augmentation and pre-processing techniques were employed to get the best performing model achieving an AUC of 0.99. Precision and recall of the model were 100 and 97.3%. An AG-CNN model was developed to recognize thorax disease. Chest X-ray 14 data set was used to detect thorax disease from X-ray images. Global and local branch attention-guided CNN was used for classification purposes. Their model was better than other models mentioned in their research

paper, achieving an AUC of 0.868. A deep convolutional neural network model was developed to classify chest X-ray images into pneumonia and other 14 dataset was used for training the model.

[11] R. Bhubaneswar et al. use the Native Bayes classifier for medical use. The authors used two well-known algorithms, the Back Propagation Neural Network (BNN) and the Native Bayesian (NB) data mining classification to study the previous experience and to calculate the probability of an object among all objects. Bayesian techniques have been developed for probability concepts. The previous back end is computed by bay rules based on the exact nature of the probability model and the Native Bayes classifier is used to study very efficiently in the supervised learning environment.

Here in this project we will be using the most widely used KNN, CNN ANN, Random Forest and Decision Tree algorithm to design a patient-oriented system for prediction of disease and to suggest a doctor based on that. This system will build the real time communication in between patient and doctors which is rarely in practice in running years. Besides, we also improve the efficiency of the algorithm.

3.REQUIREMENT ANALYSIS

3.1 SOFTWARE SPECIFICATION

3.1.1 PYTHON

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Python supports modules and packages, which encourages program modularity and code reuse. We have used Django which is one of the popular web frameworks of python that encourages rapid application development with its build in libraries.

3.1.2 SKLEARN

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

3.1.3 PANDAS

Pandas is a python library that is used to read different csv files and merge csv files. It is used for different preprocessing which includes cleaning the dataset and transferring the datasets used. Finally working data set is generated. Pandas is used in the project for exploratory data analysis.

3.1.4 NUMPY

Numpy is used for working with arrays and multidimensional matrix in the project. The project includes dimensionality reduction and generating vector space and for this numpy is used.

3.1.5 SEABORN AND MATPLOTLIB

Seaborn and matplotlib is a python library that is used for enhanced data visualization and communication in general of different data sets used in our project. We have used several plots like line plot, bar plot, scatter plot and histogram for visualizing the data sets which helps in exploration and understanding of data.

3.1.6 DJANGO

Django is a high-level Python Web framework that encourages rapid development and clean, pragmatic design. Built by experienced developers, it takes care of much of the hassle of Web development, so it can focus on writing your app without needing to reinvent the wheel. It’s free and open source. Django helps to take applications from concept to completion as quickly as possible. Django takes security seriously and helps to avoid many common security mistakes.

3.1.7 MYSQL

MySQL is developed, distributed, and supported by Oracle Corporation. MySQL is a database system used on the web it runs on a server. MySQL is ideal for both small and large applications. It is very fast, reliable, and easy to use. It supports standard SQL.

MySQL can be compiled on a number of platforms. The data in MySQL is stored in tables. A table is a collection of related data, and it consists of columns and rows. Databases are useful when storing information categorically. It is much easier and faster to manipulate data in MySQL.

3.1.8 HTML

It is the standard markup language used to create web Pages in our application. HTML is written in the form of HTML elements consisting of tags enclosed in angle brackets (e.g. <html>). HTML is skeleton of the system.

3.1.9 CASCADING STYLE SHEET(CSS)

It is a style sheet language used for describing the look and formatting of a document written in a markup language. While most often used to style web pages and interfaces written in HTML the language can be applied to any kind of XML document. CSS is used to make the User Interface more attractive as well as to add responsiveness to our system so that the system is much more comfortable to use.

3.1.10 JAVASCRIPT

JavaScript is the scripting language of the Web used in our project to make Html pages more dynamic and interactive.

3.1.11 BOOTSTRAP

Bootstrap is a free and open source front end development framework for the creation of websites and web apps. The Bootstrap framework is built on HTML, CSS and JavaScript to facilitate the development of responsive, mobile-first sites and apps. Responsive design makes it possible for a web page or app to detect the visitors screen size and orientation and automatically adapt the display accordingly in our application.

3.2 FUNCTIONAL REQUIREMENT

Functional requirements may involve calculations, technical details, data manipulation and processing, and other specific functionality that define what a system is supposed to accomplish. Various functional requirements of our system are as follows:

- Create new and valid account.
- Valid the login credentials.
- Patient should upload x-ray image for pneumonia prediction.
- Patient should fill up heart and diabetes report parameter to check problem.
- Patient should choose symptoms for disease prediction.
- Display the list of doctors.
- Doctor should manage appointment.
- Patient must take appointment to consult with doctor in real.
- Doctor must have to enter valid email for sending prescription.

3.3 NON-FUNCTIONAL REQUIREMENT

The non-functional requirements like performance, information, economy, control and security efficiency and services are very essential for successful project completion. The non-functional requirements of the project are as follows:

- The system should be easy to use, user friendly in operation.
- The system should perform with efficient throughput and response time.
- The system should provide good accuracy.

4. PROJECT DETAILS AND UML DIAGRAMS

Designing according to the requirement specification, we have tried to make sure that the system design actually confirms the user requirements of the system.

4.1 USE CASE DIAGRAM

A use case diagram is a dynamic or behavior diagram in UML. Use case diagrams model the functionality of a system using actors and use cases. Use cases are a set of actions that the system needs to perform. Use case is a list of steps, typically defining interaction between a role and a system, to achieve a goal. A use case diagram is a graphic depiction of the interactions among the elements of a system. In this context, the term "system" refers to something being developed or operated. Use case diagrams are employed in UML (Unified Modeling Language), a standard notation for the modeling of real-world objects and systems. The actors for our system are: Admin and patient and doctor. The graphical representation of what our system must actually do is represented as in use case below.

In Use Case Diagram, new patient can register into the system. Patient has additional features to login and perform heart check, pneumonia check and disease prediction. Patient can view list of doctors as well as search for a particular doctor. Patient can search for a favorable appointment schedule while booking appointment and can also manage appointment (view details, cancel appointment) after being authenticated to system. Patient can browse his profile and can edit his profile. Patient can also provide feedback regarding the system.

Doctor has additional features to login into the system and views list of patients. Doctor can manage appointments (add, edit, delete, cancel) appointment. Doctor can send prescription to particular email. On the other hand, admin can manage doctor (add, edit, delete and search doctor). Admin can also manage patients (view, edit, delete and search doctor). Admin can view the feedback provided by patient and can also assign doctor to take charge of particular disease.

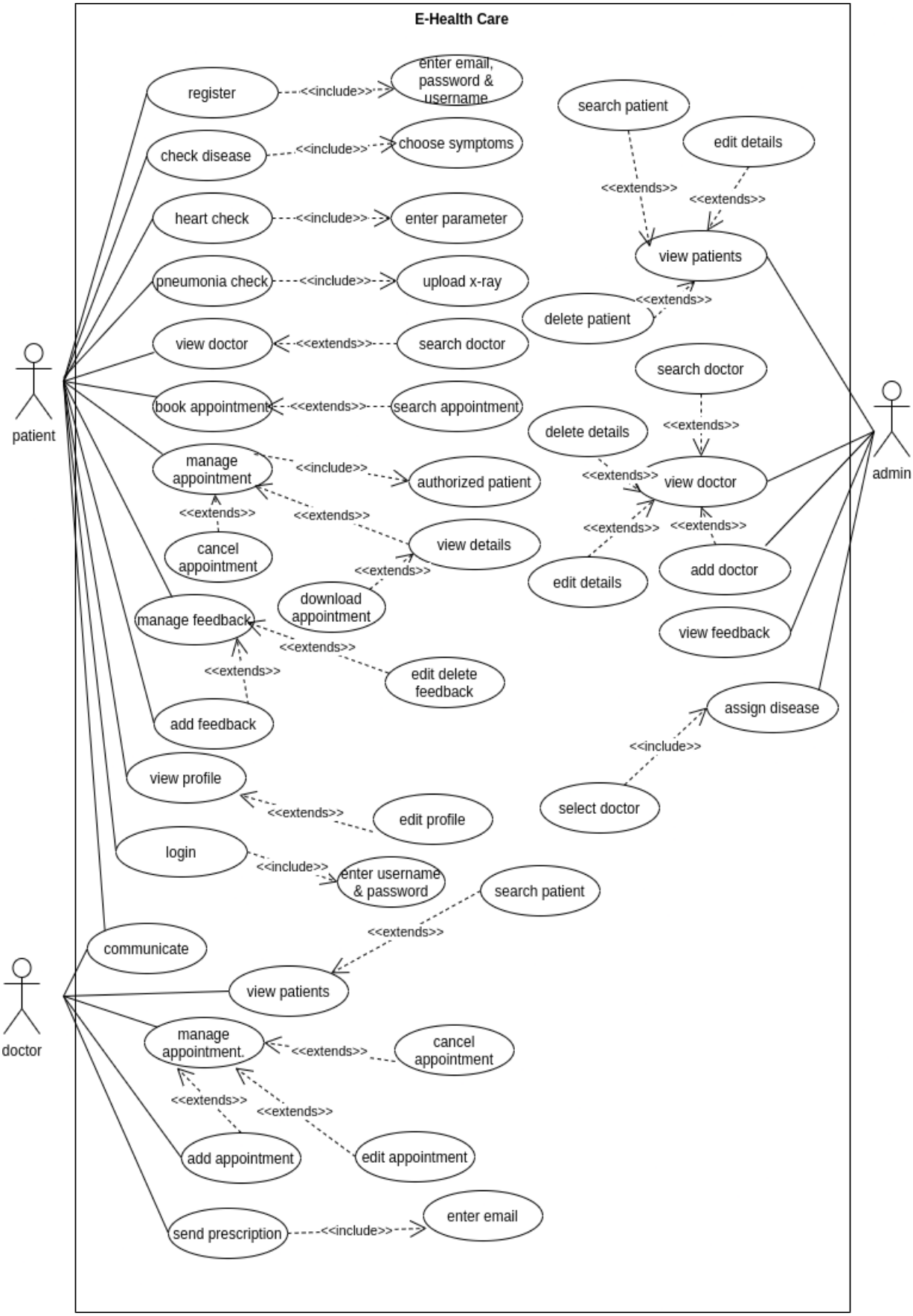

Figure 1 : Use case diagram

4.2 CONTEXT DIAGRAM

A context diagram gives an overview and it is the highest level in a data flow diagram, containing only one process representing the entire system. It should be split into major processes which give greater detail and each major process may further split to give more detail.

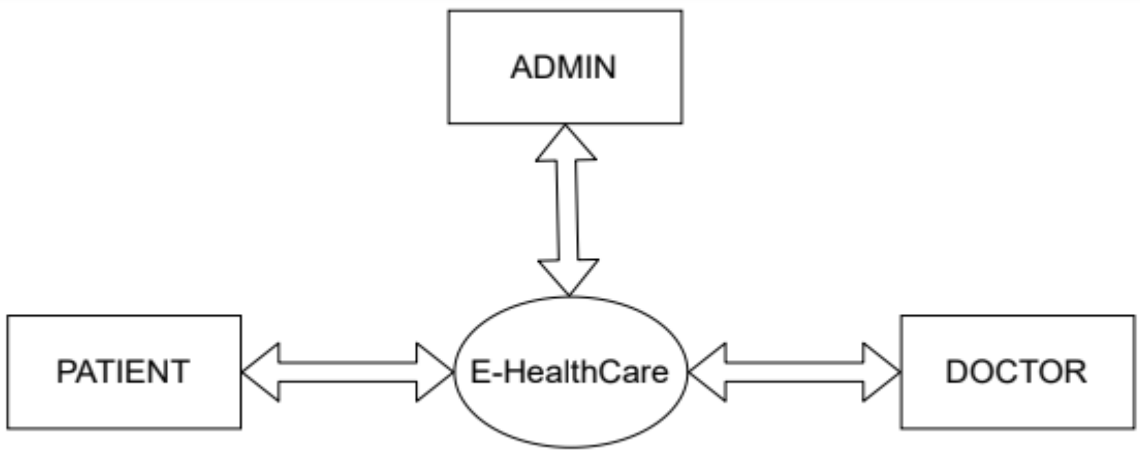

Figure 2 : Context Diagram

4.3 CLASS DIAGRAM

A class diagram is a type of diagram and part of a unified modeling language (UML) that defines and provides the overview and structure of a system in terms of classes, attributes and methods, and the relationships between different classes. It is used to illustrate and create a functional diagram of the system classes and serves as a system development resource within the software development life cycle. It is a type of structure diagram and looks similar to a flow chart having three main parts illustrated in rectangular boxes. The first or top part specifies the class name, the second or middle specifies attributes of that class and the third or bottom section lists the methods or operations that specific class can perform.

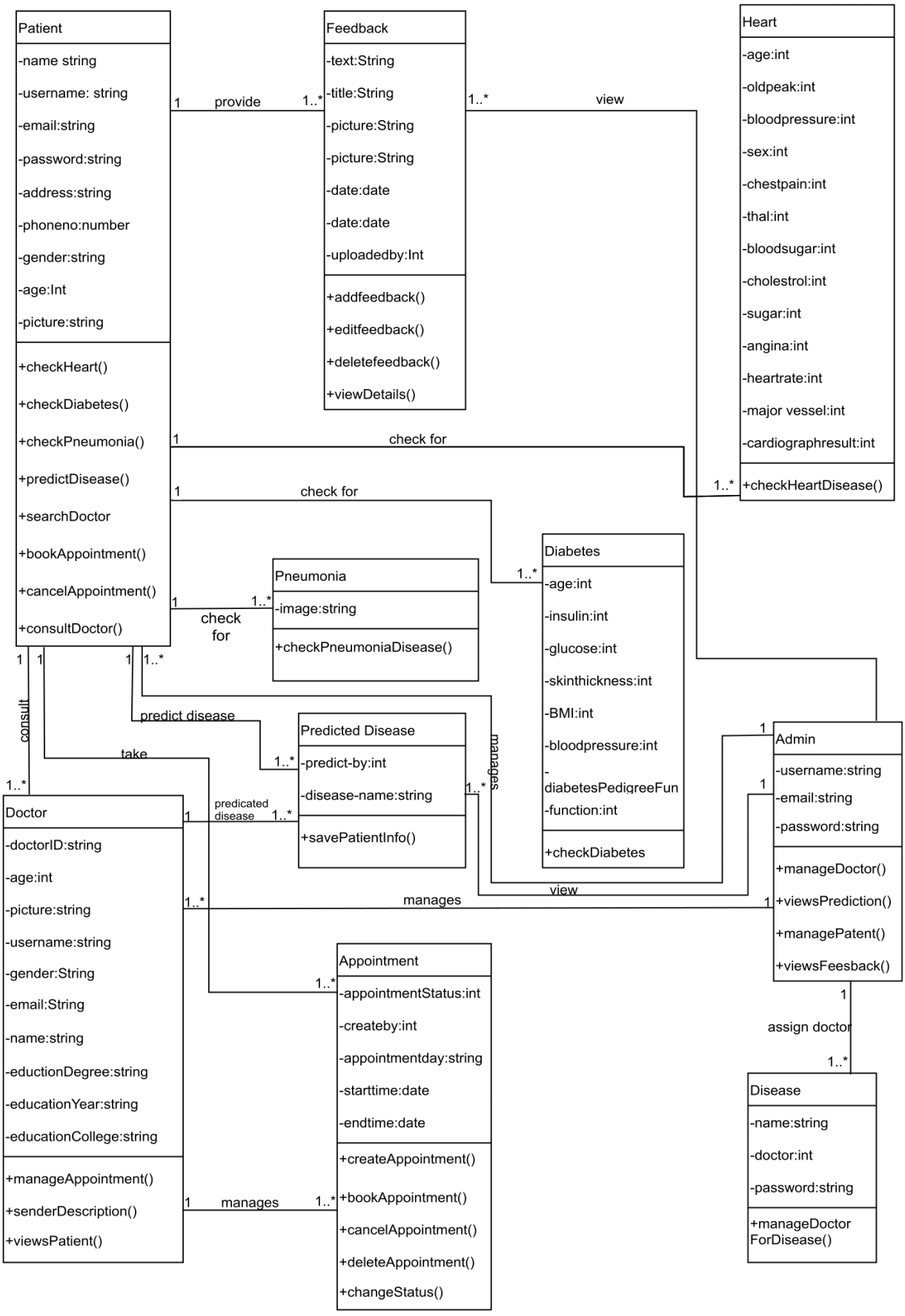

Figure 3 : Class Diagram

This project class diagram has ten different classes in which each of the classes are divided into three section. Top section includes name of the class. The second one is used to show the attributes and final one is used to describe the operations performed by the classes. Patient class have one to one or many relationships with heart class as a single patient can check multiple times for heart disease. Similarly, patient class have one to one or many relationships with pneumonia, diabetes, predicted disease class respectively. Single patient can provide one or multiple feedback regarding the system so there exists one to one or many relationships between patient class and feedback class. Patient class have one to one or many relationships with class doctor as single patient can consult with multiple doctors. Patient class also have one to one or many relationships between appointment class as single patient can book one or multiple appointment. Similar kind of relationship holds between doctor class and appointment class as on doctor can manage one or more appointment. Admin and doctor class also holds one to one or many relationships as one admin can manages one or multiple doctor. Also, admin and disease class have one to one or many relationships as admin can assign a doctor to multiple disease.

4.4 ENTITY RELATIONSHIP DIAGRAM

This depicts the relationship between the data objects. The attribute of each data objects noted in the entity relationship diagram cam be described using a data object description. Data flow diagram primarily serves two purposes.

- To provide an indication of how data are transformed as they move through the system.
- To depicts the functions that plays role in data flow.

DATA OBJECTS

A data object is a representation of almost any composite information that must be understood by the software. By composite information, we mean something that has a number of different properties or attributes. A data object encapsulates data only there is no reference within a data object to operation that act on data.

ATTRIBUTES

Attributes define the properties of data object and take on one of three different characteristics. They can be used to name an instance of data object, describe the instance, refer to another instance in other table.

RELATIONSHIP

Data objects are connected to one another in variety of different ways. We can define a set of object relationship pairs that define the relevant relationship.

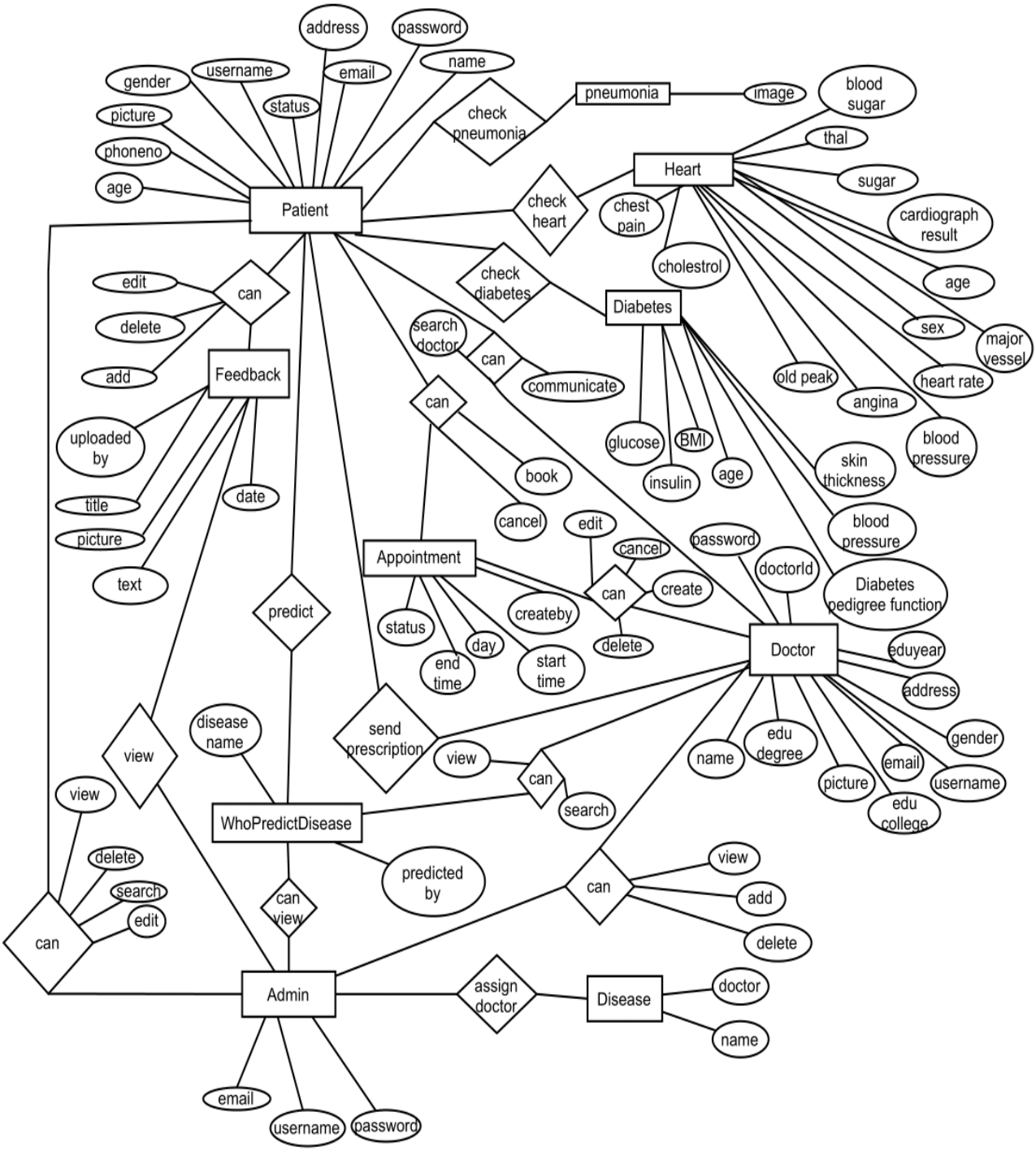

Figure 4 : ER Diagram

4.5 DATA FLOW DIAGRAM

A data flow diagram (DFD) is a graphical representation of the "flow" of data through an information system, modeling its process aspects. We used DFD as a preliminary step to create an overview of the system, which can later be elaborated also be used for the visualization of data processing (structured design) .

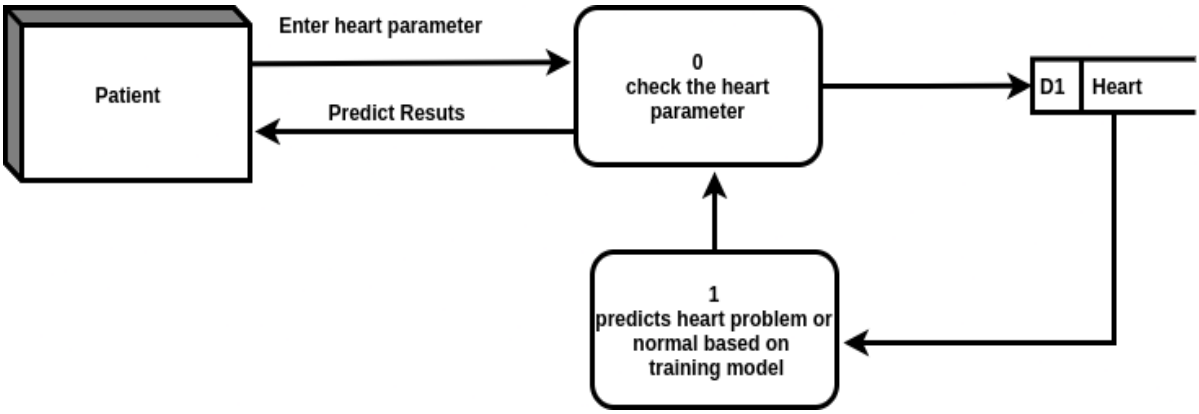

Figure 5 : DFD for heart problem prediction

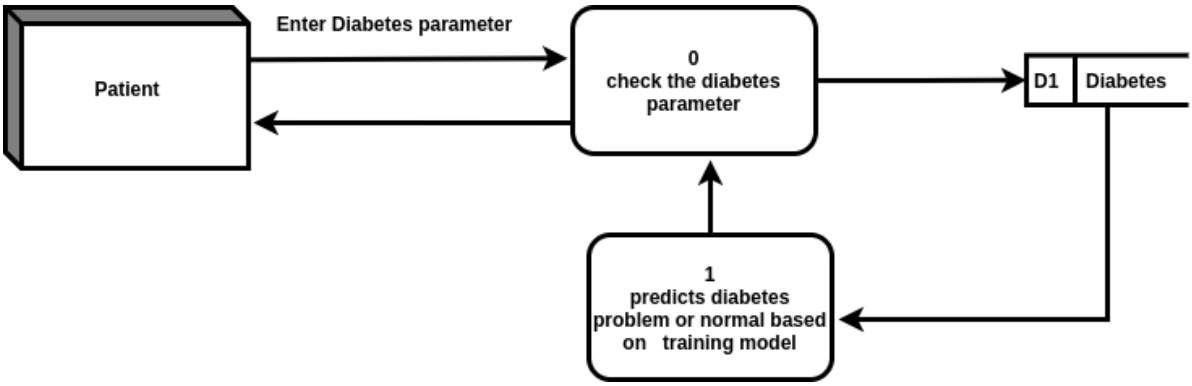

Figure 6 : DFD for diabetes problem prediction

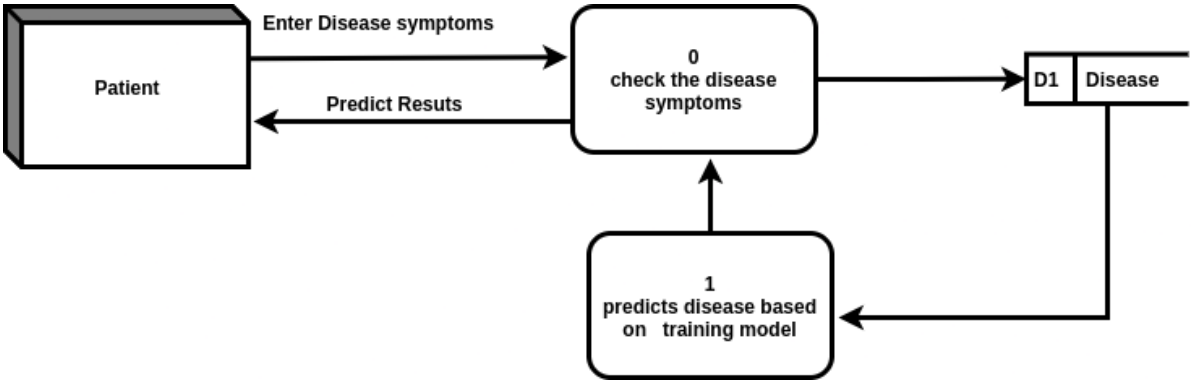

Figure 7 : DFD for disease prediction

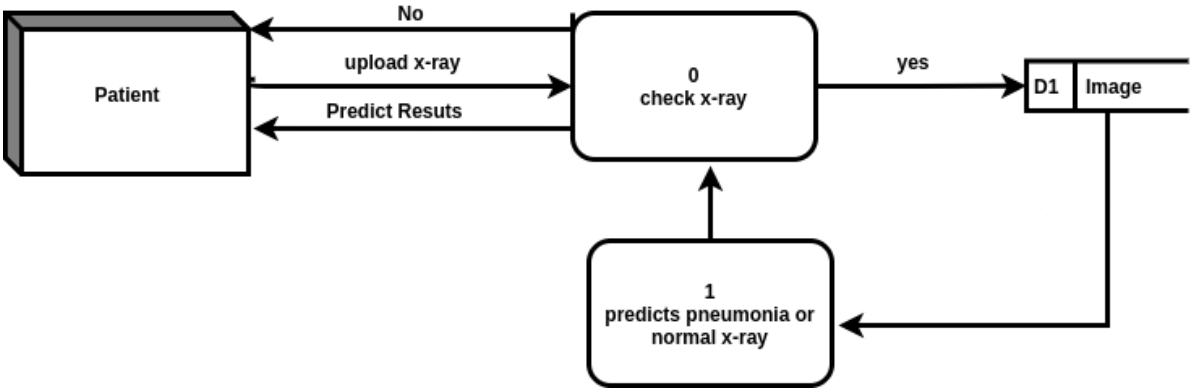

Figure 8 : DFD for pneumonia prediction

5. PROPOSED METHODOLOGY

This web application can be implemented by using Django framework of Python. Our project is based on the principle of machine learning especially based on K-nearest neighbors algorithm along with Random forest and CNN algorithm. The goal of this application is to provide health care through the form of electronic method. Technical analysis is concerned with determining how feasible a system is from a technical perspective. Python programming language will be used for backend and HTML and CSS for the frontend design. Current tools are more than enough for system development

We have planned to work following these methodologies for the application of knowledge, skills and technique to broad range of activities in order to meet the requirement of our project.

5.1 SOFTWARE DEVELOPMENT LIFECYCLE

The framework we will be using for developing this project is iterative model of software development life cycle. In this model, a simple and primitive implementation of very small set of software requirement is done at first, which is followed by the iterative enhancement in the primitive model until all requirements are fulfilled and the software is ready for deployed. The following sub section briefly describe various phase in iterative model of SDLC that was applied in the development of system.

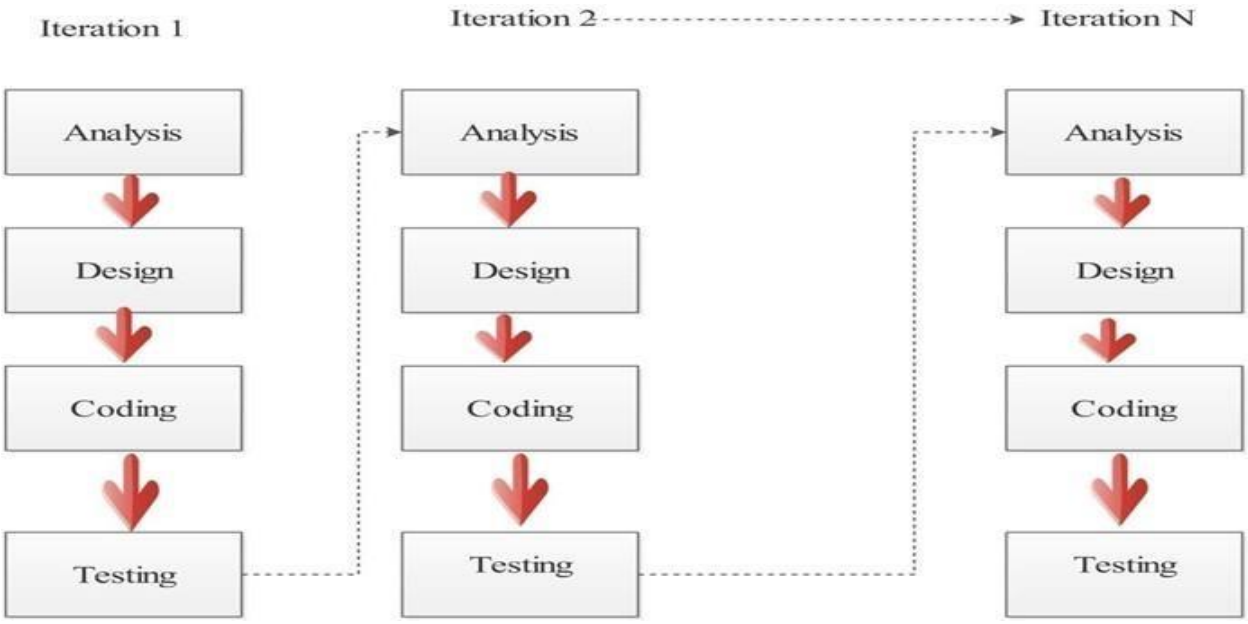

Figure 9 : Iterative model of software development life cycle

5.1.1 REQUIREMENT ANALYSIS

In this phase, analysis will be performed in order to find out the requirements of the system. The outcome of this phase would be a SRS which is acronym for "system requirement specifications"

5.1.2 DESIGN PHASE

In this phase the SRS would be translated into the system design. Context diagram DFD ED diagram, use case diagram and class diagram will be developed.

5.1.3 CODING PHASE

In this phase coding will be done according for the design and a working system will be developed by the end of the process.

5.1.4 TESTING PHASE

In this phase the system will be tested with each testing list of changes to the system developed, is suggested and the change will be applied to the software and the software would be delivered as a successive increment until a satisfying system is achieved.

5.2 EXPLORATORY DATA ANALYSIS

DATASET DESCRIPTION

- **Heart.csv:** This file contains data of heart report like age, sex, chest pain, cholesterol, resting blood pressure (trestbps), serum cholesterol in mg/dl (chol), fasting blood sugar > 120 mg/dl (fbs) , resting electrocardiographic results (restecg), maximum heart rate achieved (thalach), exercise induced angina (exang) , ST depression induced by exercise relative to rest(old peak), the slope of the peak exercise ST segment(slope), number of major vessels (0-3) colored by fluoroscopy(ca) ,thal (thallium stress result). The datasets contain sample of 1025 data. The above-mentioned data are input data which are independent to each other but the target data is predicted value which is dependent on input value of given data sets. Target value contains 0 and 1 as a output which denotes no heart problem prediction and heart disease problem respectively.

Exploration of heart.csv

In our data sets, various data are represented in integer and floating-point value and it does not include null value in any of data which are depicted by figure below:

```
[11]: df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1025 entries, 0 to 1024
Data columns (total 14 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   age       1025 non-null   int64
 1   sex       1025 non-null   int64
 2   cp        1025 non-null   int64
 3   trestbps  1025 non-null   int64
 4   chol      1025 non-null   int64
 5   fbs       1025 non-null   int64
 6   restecg   1025 non-null   int64
 7   thalach   1025 non-null   int64
 8   exang     1025 non-null   int64
 9   oldpeak   1025 non-null   float64
 10  slope     1025 non-null   int64
 11  ca        1025 non-null   int64
 12  thal      1025 non-null   int64
 13  target    1025 non-null   int64
dtypes: float64(1), int64(13)
```

Figure 10 : Information about heart.csv

As we already mentioned above that target value consists 0 and 1 in the datasets. The total number of 0 value count in datasets is 499 which indicates the number of normal heart report data and number of 1 value count is 526 which indicates the number of data suffering from heart problem. The above mention information is depicted graphically in the figure below.

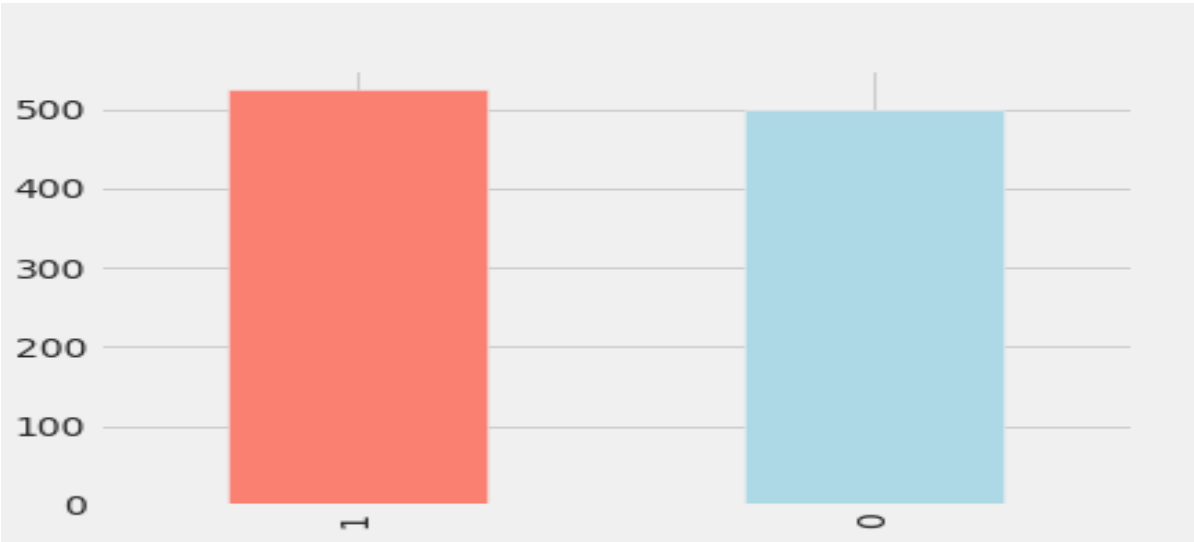

Figure 11: Value count for normal condition and heart problem

The above datasets do not contain missing value of any data presented in the file. The output combination of dependent and independent data is described and represented graphically in figure below:

cp (Chest Pain): People with cp equal to 1, 2, 3 are more likely to have heart disease than people with cp equal to 0.

restecg (resting electrocardiographic results): People with value 1 (signals non-normal heart beat, can range from mild symptoms to severe problems) are more likely to have heart disease.

exang (exercise induced angina): People with value 0 (no exercise induced angina) have heart disease more than people with value 1 (yes exercise induced angina)

slope (the slope of the peak exercise ST segment): People with slope value equal to 2 (down sloping, signs of unhealthy heart) are more likely to have heart disease than people with slope value equal to 0 (up sloping, better heart rate with exercise) or 1 (flat sloping, minimal change (typical healthy heart)).

ca (number of major vessels (0-3) colored by fluoroscopy): the more blood movement the better so people with ca equal to 0 are more likely to have heart disease.

thal (thallium stress result) : People with that value equal to 2 (fixed defect: used to be defect but ok now) are more likely to have heart disease.

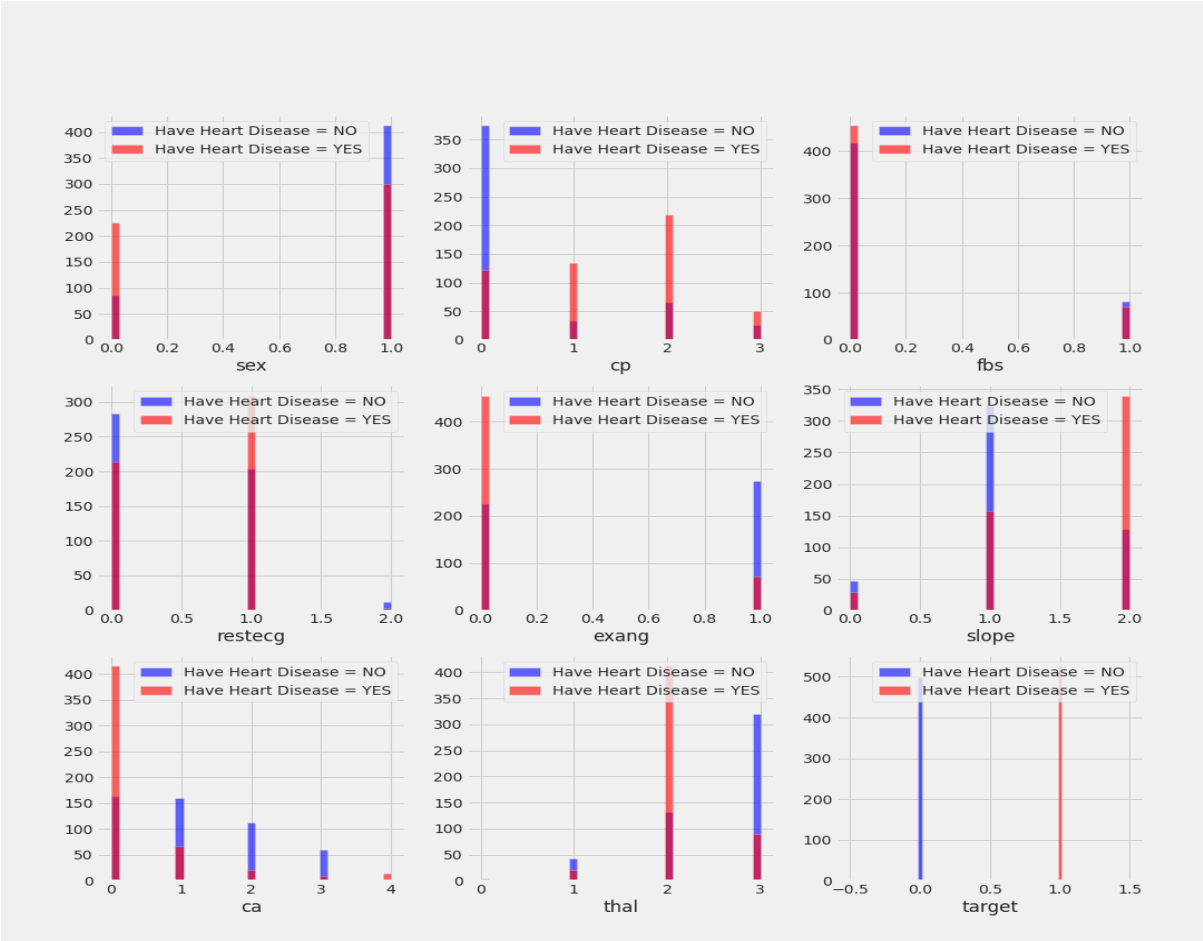

Figure 12: Heart problem analysis with respect to independent variable

The figure below shows the relationship between age and maximum heart rate achieved. We have explored that maximum heart rate is achieved in people of age from 40 to 70 years so it is considered as the age in which maximum heart problem occurred. This is depicted graphically in figure below:

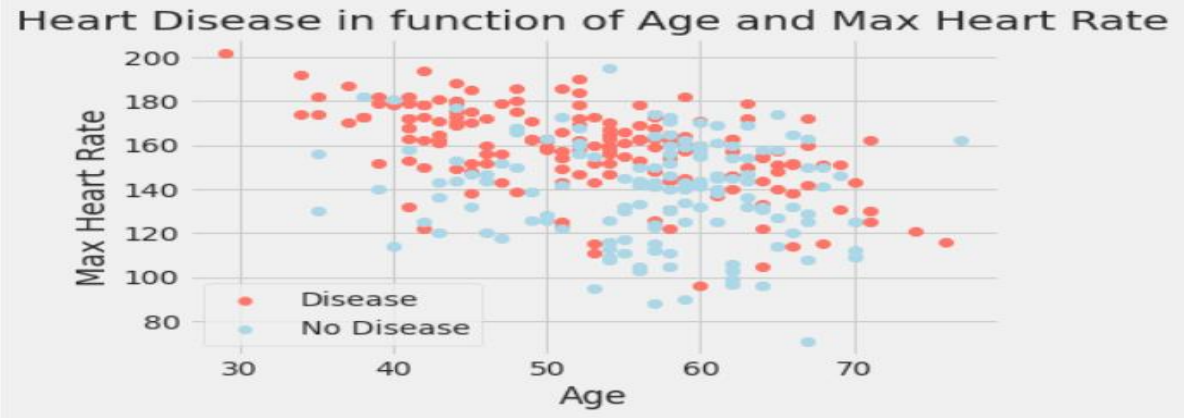

Figure 13: Heart disease in function of age and max heart rate

Data processing of heart.csv

After exploring the dataset, we observed that we need to convert some categorical variables into dummy variables and scale all the values before training the machine learning models. First, we will use the get dummies method of pandas library to create dummy columns for categorical variables.

```
In [18]: dataset.head()
```

Out[18]:

	age	trestbps	chol	thalach	oldpeak	target	sex_0	sex_1	cp_0	cp_1	...	slope_2	ca_0
0	52	125	212	168	1.00	0	0	1	1	0	...	1	0
1	53	140	203	155	3.10	0	0	1	1	0	...	0	1
2	70	145	174	125	2.60	0	0	1	1	0	...	0	1
3	61	148	203	161	0.00	0	0	1	1	0	...	1	0
4	62	138	294	106	1.90	0	1	0	1	0	...	0	0

5 rows × 31 columns

Figure 14 : Data Processing of heart.csv

5.3 ALGORITHM USED

5.3.1 K-NEAREST NEIGHBOUR (KNN) ALGORITHM

- K-Nearest Neighbor is one of the simplest Machine Learning algorithms based on Supervised Learning technique.
- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.
- K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.
- K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.

We have used K-NN algorithm for prediction of heart disease. The K-NN working can be explained on the basis of the below algorithm:

Step-1: Select the number K of the neighbors

Step-2: Calculate the Euclidean distance of K number of neighbors

Step-3: Take the K nearest neighbors as per the calculated Euclidean distance.

Step-4: Among these k neighbors, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbor is maximum. Finally, our model is ready.

We have a new data point and we need to put it in the required category. Consider the below image:

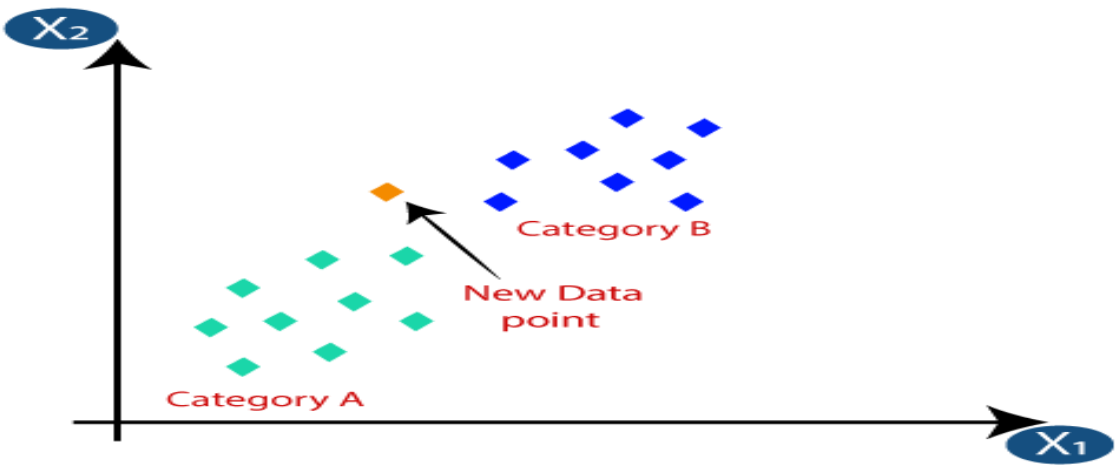

Figure 15 : New data point and category

Firstly, we will choose the number of neighbors, so we will choose the k=5.

Next, we will calculate the Euclidean distance between the data points. The Euclidean distance is the distance between two points. It can be calculated as seen in the image below:

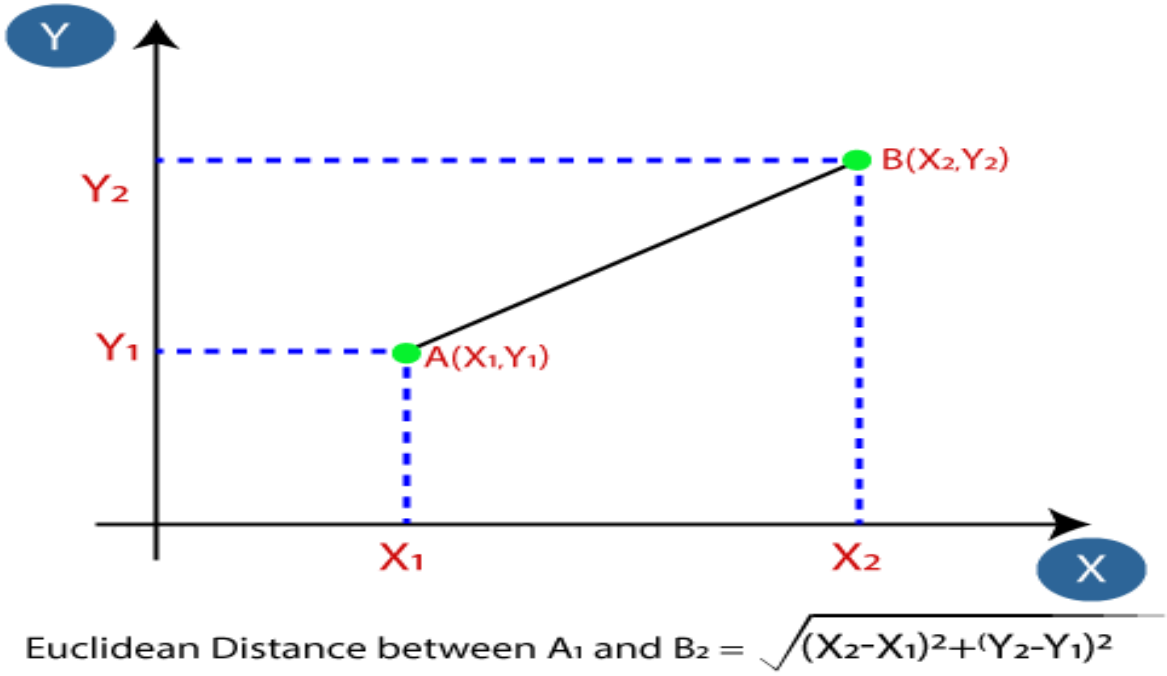

Figure 16 : Calculation of Euclidean distance between two points

By calculating the Euclidean distance, we got the nearest neighbors, as three nearest neighbors in category A and two nearest neighbors in category B. Consider the below image:

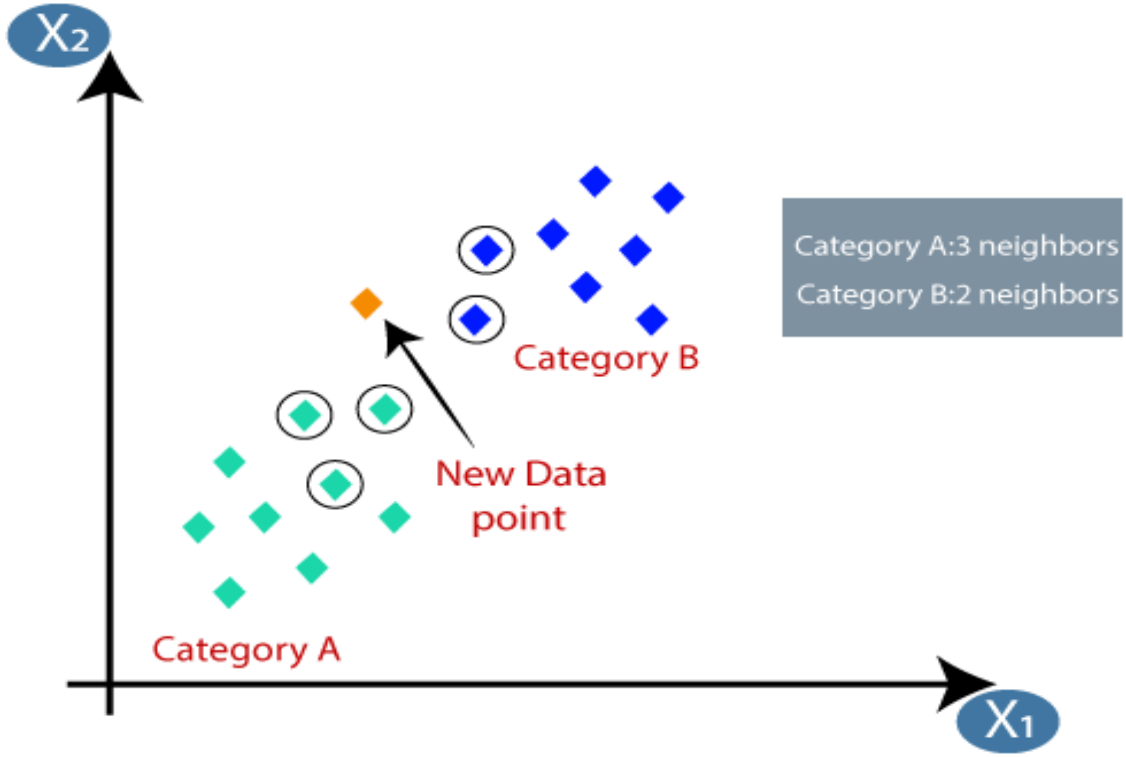

Figure 17 : Selection of category for new data point

As we can see the 3 nearest neighbors are from category A, hence this new data point must belong to category A.

Below are some points to remember while selecting the value of K in the K-NN algorithm:

There is no particular way to determine the best value for "K", so we need to try some values to find the best out of them. The most preferred value for K is 5.

A very low value for K such as K=1 or K=2, can be noisy and lead to the effects of outliers in the model.

Large values for K are good, but it may find some difficulties.

K-NN ACCURACY FOR HEART DISEASE DREDICTION

We have split the datasets into two parts that is for training size we have used 75 percent of the data and 25 percent is used for testing size. The accuracy score of algorithms after training is 91.77 percent and accuracy score for testing new data is 81.82 percent.

Train results of confusion matrix

1.310 is the amount of true Positives in our data, while 348 is the amount of true negatives.

2.29 and 30 are the number of errors.

3. There are 30 false positives error that is algorithm have predicted positive but the actual result is false.

4. There are 29 false negatives that is algorithm have predicted negative and it is false.

5. Hence, if we calculate the accuracy its Correct Predicted/Total. In other words, where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives.

Accuracy = (TP + TN)/(TP + TN + FP + FN).

Accuracy = (310+348)/(348+310+29+30) = 0.80 = 91.77% accuracy

Test results of confusion matrix

1.123 is the number of true Positives in our data, while 129 is the number of true negatives.

2.20 and 36 are the number of errors.

3. There are 36 false positives error that is algorithm have predicted positive but the actual result is false.

4. There are 20 false negatives that is algorithm have predicted negative and it is false.

5. Hence, if we calculate the accuracy its Correct Predicted/Total. In other words, where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives.

Accuracy = (TP + TN)/(TP + TN + FP + FN).

Accuracy =(123+129)/(123+129+36+20) = 81.82% accuracy

```
Train Result:
================================================
Accuracy Score: 91.77%
________________________________________________
CLASSIFICATION REPORT:
                    0       1   accuracy  macro avg  weighted avg
precision        0.91    0.92       0.92       0.92          0.92
recall           0.91    0.92       0.92       0.92          0.92
f1-score         0.91    0.92       0.92       0.92          0.92
support        340.00  377.00       0.92     717.00        717.00
________________________________________________
Confusion Matrix:
 [[310  30]
 [ 29 348]]

Test Result:
================================================
Accuracy Score: 81.82%
________________________________________________
CLASSIFICATION REPORT:
                    0       1   accuracy  macro avg  weighted avg
precision        0.86    0.78       0.82       0.82          0.82
recall           0.77    0.87       0.82       0.82          0.82
f1-score         0.81    0.82       0.82       0.82          0.82
support        159.00  149.00       0.82     308.00        308.00
________________________________________________
Confusion Matrix:
 [[123  36]
 [ 20 129]]
```

Figure 18 : K-NN accuracy for heart problem prediction

- **Diabetes.csv:** This file contains data of diabetes report like pregnancies, glucose, blood pressure, skin thickness, insulin, bmi, diabetes pedigree function, age outcome the datasets contain sample of 1025 data. The above-mentioned data are input data which are independent to each other but the outcome data is predicted value which is dependent on input value of given data sets. Outcome value contains 0 and 1 as a output which denotes no diabetes problem and diabetes disease problem respectively.

Exploration of diabetes.csv

In our data sets, various data are represented in integer and floating-point value and it does not include null value in any of data which are depicted by figure below:

```
[5]: diabetes_data.info(verbose=True)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2000 entries, 0 to 1999
Data columns (total 9 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Pregnancies               2000 non-null   int64
 1   Glucose                   2000 non-null   int64
 2   BloodPressure             2000 non-null   int64
 3   SkinThickness             2000 non-null   int64
 4   Insulin                   2000 non-null   int64
 5   BMI                       2000 non-null   float64
 6   DiabetesPedigreeFunction  2000 non-null   float64
 7   Age                       2000 non-null   int64
 8   Outcome                   2000 non-null   int64
dtypes: float64(2), int64(7)
memory usage: 140.8 KB
```

Figure 19 : Information about diabetes.csv

As we already mentioned above that outcome value consists 0 and 1 in the datasets. The total number of 0 value count in datasets is 1316 which indicates the number of normal diabetes report data and number of 1 value count is 684 which indicates the number of data suffering from diabetes problem. The above mention information is depicted graphically in the figure below.

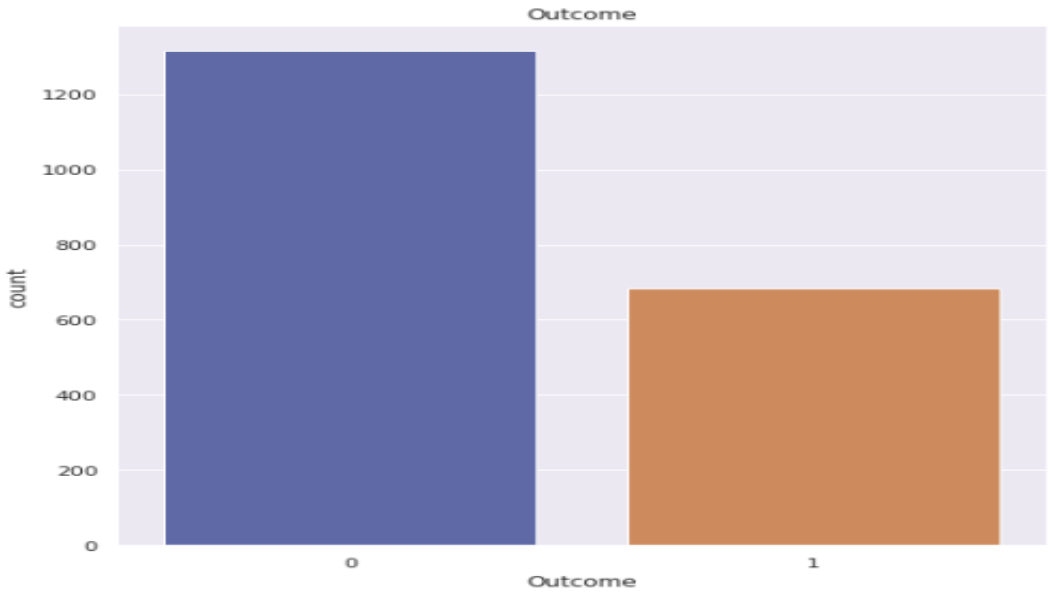

Figure 20 : Value count for normal condition and diabetes problem

The figure below shows the relationship between age and maximum diabetes rate achieved. We have explored that maximum diabetes rate is achieved in people of age from 40 to 70 years so it is considered as the age in which maximum diabetes problem occurred. This is depicted graphically in figure below:

Figure 21 : Diabetes disease in function of age and maximum diabetes rate

The above datasets do not contain missing value of any data presented in the file. The figure below depicted total number of counts according to each individual parameter present in the datasets.

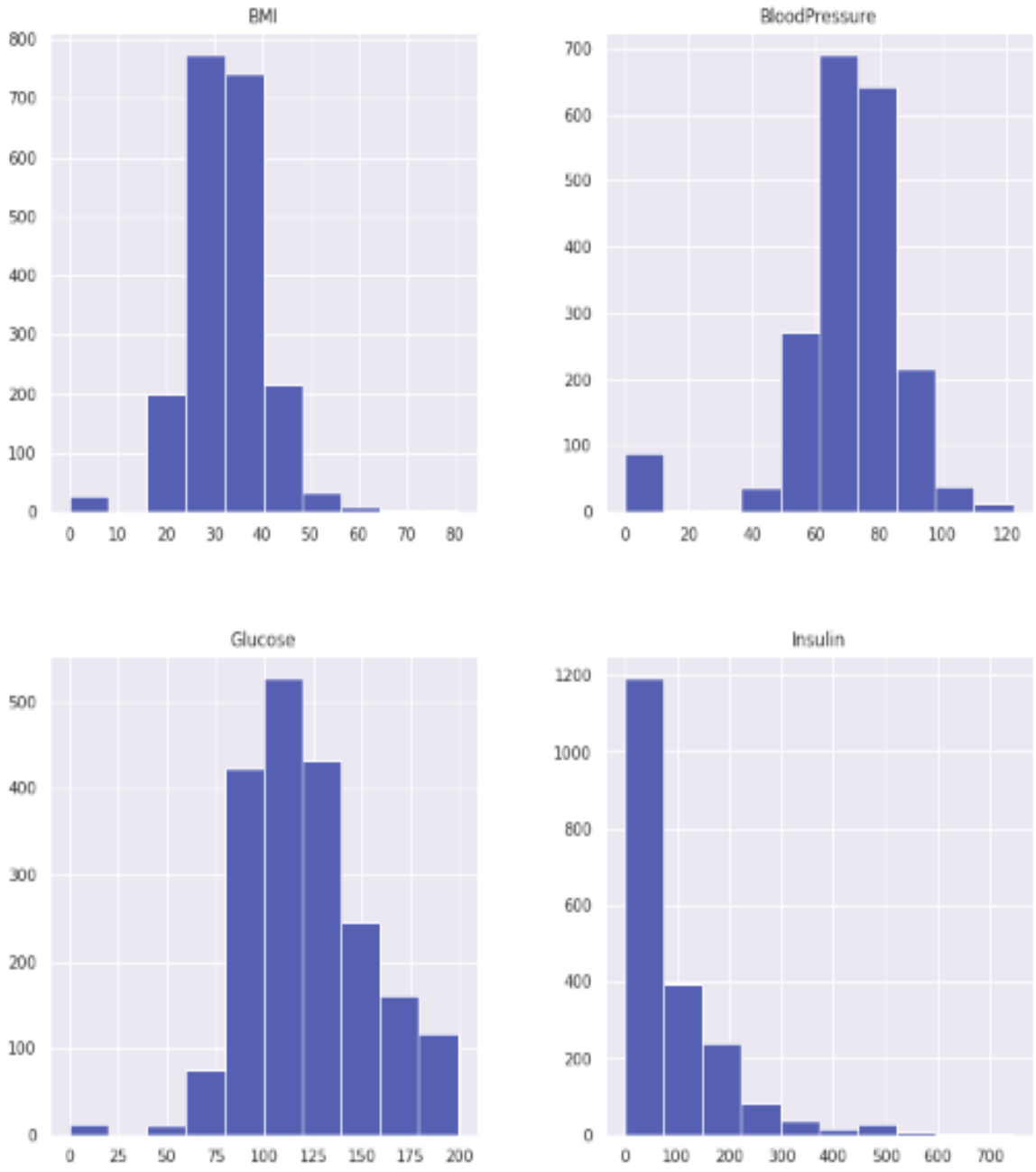

Figure 22 : Total count of individual data

K-NN ACCURACY FOR DIABETES DISEASE DREDICTION

We have split the datasets into two parts that is for training size we have used 75 percent of the data and 25 percent is used for testing size. The accuracy score of algorithms after training is 88.40 percent and accuracy score for testing new data is 81.20 percent.

Train results of confusion matrix

1.887 is the amount of true Positives in our data, while 439 is the amount of true negatives.

2.79 and 95 are the number of errors.

3. There are 95 false positives error that is algorithm have predicted positive but the actual result is false.

4. There are 79 false negatives that is algorithm have predicted negative and it is false.

5. Hence, if we calculate the accuracy its Correct Predicted/Total. In other words, where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives.

Accuracy = (TP + TN)/(TP + TN + FP + FN).

Accuracy =(887+439)/(887+439+95+79) = 0.80 = 88.40% accuracy

Test results of confusion matrix

1.293 is the number of true Positives in our data, while 113 is the number of true negatives.

2.53 and 41 are the number of errors.

3. There are 41 false positives error that is algorithm have predicted positive but the actual result is false.

4. There are 53 false negatives that is algorithm have predicted negative and it is false.

5. Hence, if we calculate the accuracy its Correct Predicted/Total. In other words, where TP, FN, FP and TN represent the number of true positives, false negatives, false positives and true negatives.

Accuracy = (TP + TN)/(TP + TN + FP + FN).

Accuracy =(293+113)/(293+113+41+53) = 81.20% accuracy

```
Train Result:
================================================
Accuracy Score: 88.40%
------------------------------------------------
CLASSIFICATION REPORT:
                    0            1    accuracy     macro avg  weighted avg
precision    0.918219     0.822097       0.884      0.870158      0.885025
recall       0.903259     0.847490       0.884      0.875375      0.884000
f1-score     0.910678     0.834601       0.884      0.872639      0.884406
support    982.000000   518.000000       0.884   1500.000000   1500.000000
------------------------------------------------
Confusion Matrix:
 [[887  95]
 [ 79 439]]

Test Result:
================================================
Accuracy Score: 81.20%
------------------------------------------------
CLASSIFICATION REPORT:
                    0            1    accuracy    macro avg  weighted avg
precision    0.846821     0.733766       0.812     0.790294      0.809287
recall       0.877246     0.680723       0.812     0.778984      0.812000
f1-score     0.861765     0.706250       0.812     0.784007      0.810134
support    334.000000   166.000000       0.812   500.000000    500.000000
------------------------------------------------
Confusion Matrix:
 [[293  41]
 [ 53 113]]
```

Figure 23 : K-NN accuracy for diabetes disease prediction

- **Pneumonia.zip:** This file contains 2 gb images file which are categorized into two parts that consist of x-ray report images showing pneumonia and normal x-ray report images. The total counts of x-ray images showing pneumonia is 3875 and normal x-ray report 1341.

5.3.2 VISUALIZATION OF TRAINING IMAGES

We displayed some normal and pneumonia images to just have a look at how much different they look from the naked eye. The figure below depicts the sample of chest x-ray image showing pneumonia and normal x-ray report:

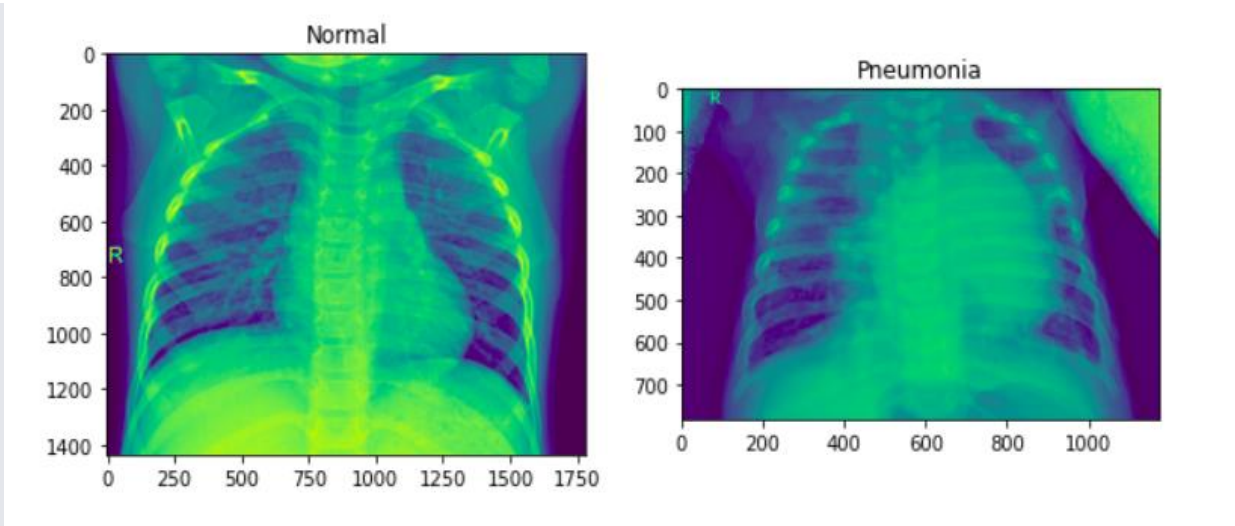

Figure 24 : Representation of chest x-ray showing pneumonia and normal report

Convolution Neural Network is one of the main categories to do image classification and image recognition in neural networks. Scene labeling, objects detections, and face recognition, etc., are some of the areas where convolution neural networks are widely used.

CNN takes an image as input, which is classified and process under a certain category such as normal x-ray, infected x-ray, and pictures other than x-ray etc. The computer sees an image as an array of pixels and depends on the resolution of the image. Based on image resolution, it will see as h * w * d, where h is height, w is width and d is dimension. For example, An RGB image is 6 * 6 * 3 array of the matrix, and the grayscale image is 4 * 4 * 1 array of the matrix.

In CNN, each input image will pass through a sequence of convolution layers along with pooling, fully connected layers, filters (Also known as kernels). After that, we will apply the Soft-max function to classify an object with probabilistic values 0 and 1.

We have used CNN for prediction of pneumonia after patient upload a x-ray image. The figure below depicts the actual working flow of the algorithm.

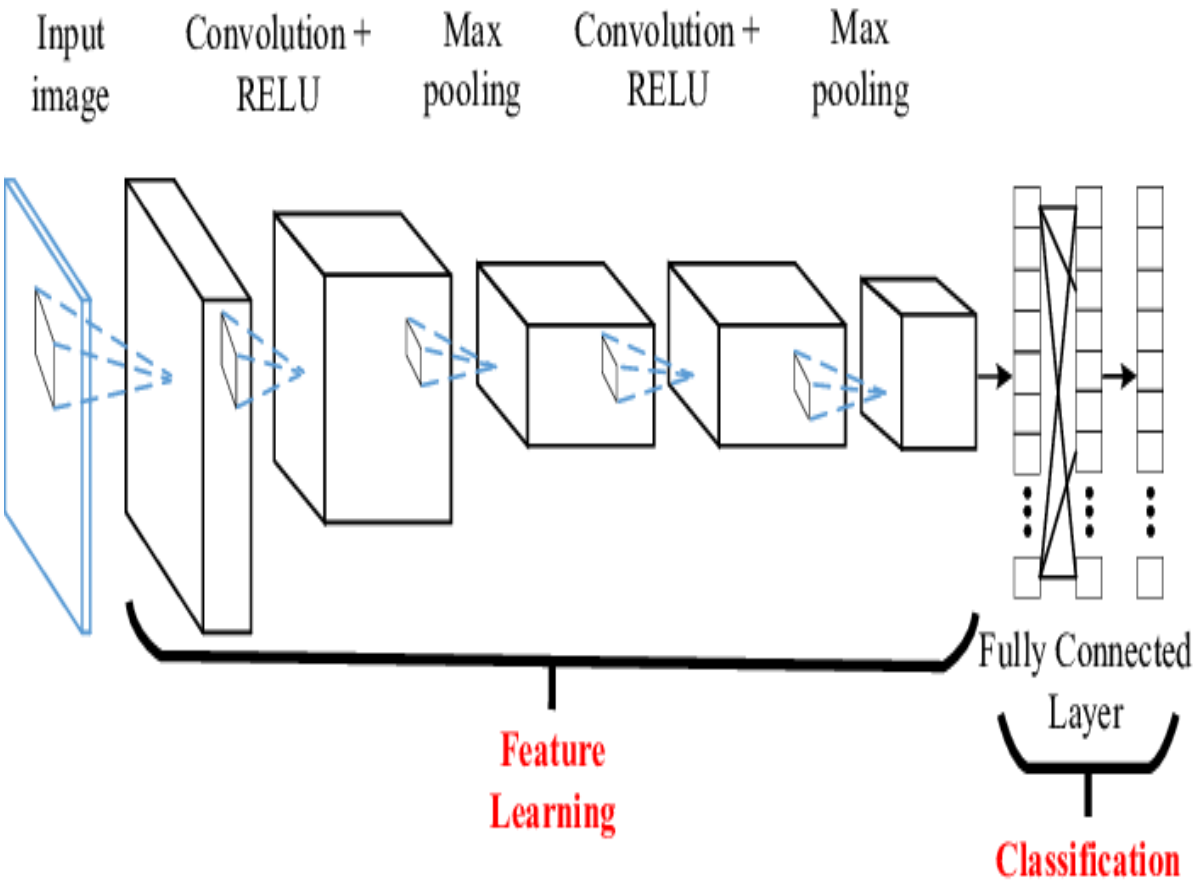

Figure 25 : Working of CNN model

Next, we have split the datasets into three different parts that is train, validation and test sets for data augmentation.

DATA AUGMENTATION

The practice of data augmentation is an effective way to increase the size of the training set. Augmenting the training examples allow the network to see more diversified, but still representative, data points during training. Then we defined a couple of data generators: one for training data, and the other for validation data. A data generator is capable of loading the required amount of data (a mini batch of images) directly from the source folder, convert them into training data (fed to the model) and training targets (a vector of attributes, the supervision signal)/ For our experiments, we usually set the batch size to 64. In general, a value between 32 and 128 should work well. Usually we should increase or decrease the batch size according to computational resources and model's performances. The next step was to build the model. This can be described in the following five steps:

- We used five convolution blocks comprised of convolution layer, max-pooling and batch-normalization.
- On top of it we used a flatten layer and followed it by four fully connected layers.
- Also in between we have used dropouts to reduce over-fitting.
- Activation function was Rely throughout except for the last layer where it was Sigmoid as this is a binary classification problem.
- We have used Adam as the optimizer and cross-entropy as the loss.

ACCURACY OF CNN ALGORITHM

The model is able to achieve a test accuracy of 91.02% which is quite good considering the size of data that is used. Also, the training accuracy of model is 98.23%.

```
CONFUSION MATRIX ------------------
[[191  43]
 [ 13 377]]

TEST METRICS ----------------------
Accuracy: 91.02564102564102%
Precision: 89.76190476190476%
Recall: 96.66666666666667%
F1-score: 93.08641975308642

TRAIN METRIC ----------------------
Train acc: 94.23
```

Figure 26 : Accuracy of CNN algorithm

- **Disease.csv:** This file contains total 4920 row of data which includes 41 diseases along with the symptoms that cause it. When six input symptoms data values are given to random forest machine leaning model it will predict the disease based on that data values. The above data sets do not contain missing value in any of the data field.

As we can see from the figure below the above data sets contains number of diseases having the same percentage through bar chart.

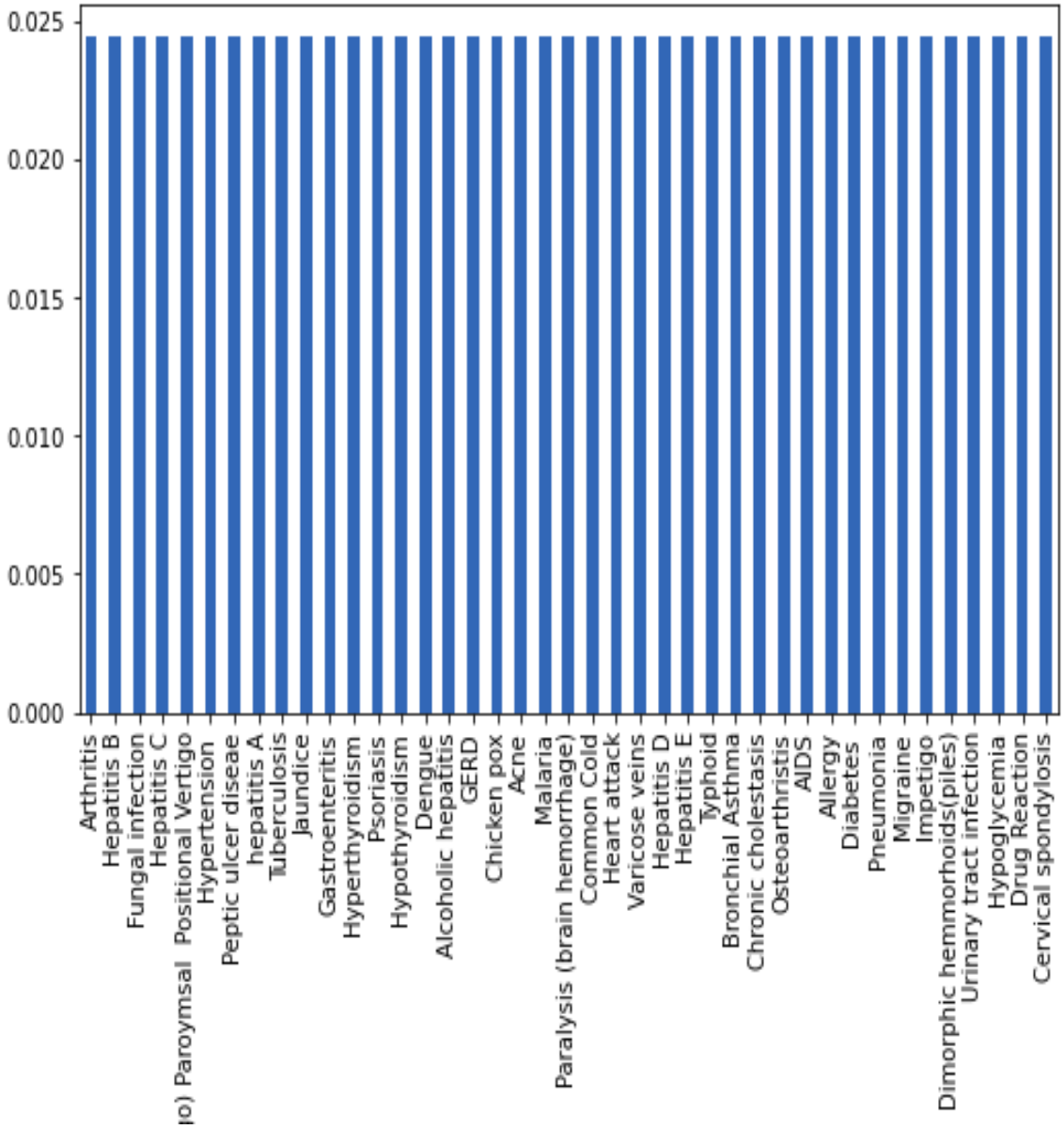

Figure 27 : Analysis of disease datasets

We have split the datasets into two parts that is for training size we have used 70 percent of the data and 30 percent is used for testing size. Next, we have applied random forest algorithm for predicting disease after selection of symptoms.

5.3.3 RANDOM FOREST ALGORITHM

Random forest is a supervised learning algorithm which is used for both classification as well as regression. But however, it is mainly used for classification problems. As we know that a forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting. It is an ensemble method which is better than a single decision tree because it reduces the over-fitting by averaging the result.

The following are the working steps of Random Forest Algorithm

Step 1 − First, we start with the selection of random samples from a given dataset.

Step 2 − Next, this algorithm will construct a decision tree for every sample. Then we will get the prediction result from every decision tree.

Step 3 − In this step, voting will be performed for every predicted result.

Step 4 − At last, we have selected the most voted prediction result as the final prediction result.

Diagram below illustrate the algorithm flow−

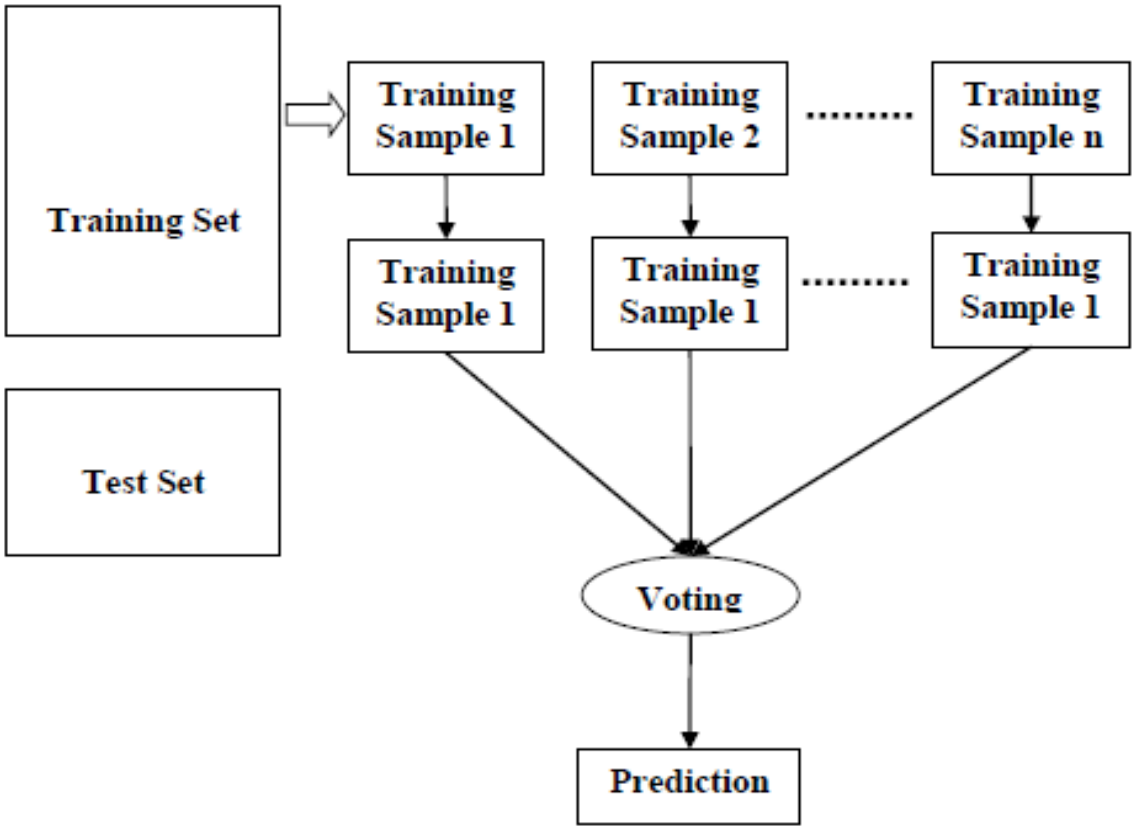

Figure 28 : Random Forest Algorithm

ACCURACY OF RANDOM FOREST ALGORITHM

After splitting the datasets into two parts that is for training size we have used 70 percent of the data and 30 percent is used for testing size and after applying it to random forest algorithm, the accuracy score of algorithms after training is 94.34 percent and accuracy score for testing new data is 93.70 percent.

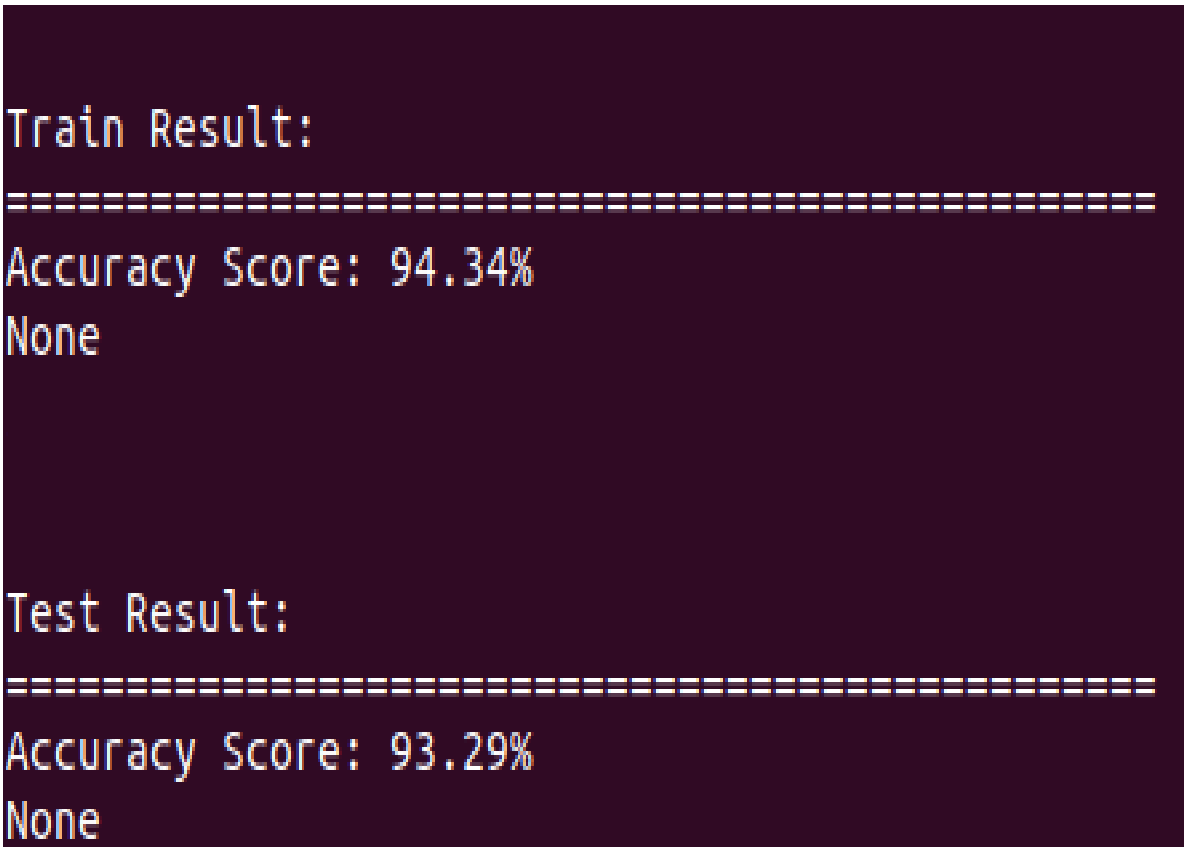

Figure 29 : Accuracy of RF algorithm

5.4 SECURING DATA IN WEB

The main problem while developing a web-based application is lack of security. All the data is accessed in the web through our device while running this application and the data must be secured .The first thing we consider while data security is physical security i.e. securing data from thieves, natural disasters and other physical harm which may damage the data and may be misused if stolen by someone .The physical protection can be provided by guards and administration's people but these people cannot secure data in the web where we should give attention.

The data in the web can be secured by following activities:

- Securing data from SQL injection. There have been many high visibility attacks against web applications that can be traced back to a SQL injection attack successfully stealing passwords. SQL injection is a code injection technique that might destroy our database. SQL injection is one of the most common web hacking techniques. SQL injection is the placement of malicious code in SQL statements, via web page input. This can be controlled by using prepared statements while coding.

- The data in web need to be secure from unauthorized users and hackers too. The data can be secured by using firewall in our device so that the unwanted file doesn't get into our device which may cause damage or loss of data.MD5 or any other standard hashing algorithms can be used to secure our login information more secure to avoid unauthorized access.

6. TESTING

Testing is advantageous in several ways. Firstly, the defects found help in the process of making the software reliable. Secondly, even if the defects found are not corrected, testing gives an idea as to how reliable the software is. Thirdly, over time, the record of defects found reveals the most common kinds of defects, which can be used for developing appropriate preventive measures such as training, proper design and reviewing.

6.1 TESTING PLAN

The testing sub-process includes the following activities in a phase dependent manner:

- Create Test Plans
- Create Test Specifications
- Review Test Plans and Test Specifications
- Conduct tests according to the Test Specifications, and log the defects
- Fix defects, if any
- When defects are fixed continue from activity

6.2 TESTING METHODS

6.2.1 BLACK BOX AND WHITE BOX TESTING

In black box testing a software item is viewed as a black box, without knowledge of its internal structure or behavior. Possible input conditions, based on the specifications (and possible sequences of input conditions), are presented as test cases. In white-box testing knowledge of internal structure and logic is exploited. Test cases are presented such that possible paths of control flow through the software item are traced. Hence more defects than black box testing is likely to be found.

TESTING PRINCIPLES

- While deciding on the focus of testing activities, study project priorities. For example, for an on-line system, pay more attention to response time. Spend more time on the features used frequently
- Decide on the effort required for testing based on the usage of the system. If the system is to be used by a large number of users, evaluate the impact on users due to a system failure before deciding on the effort
- A necessary part of the test case is a definition of the expected result
- Write test cases for invalid and unexpected as well as valid and expected input conditions
- Thoroughly inspect the results of each test

We have performed both Unit Testing and System Testing to detect and fix errors. A brief description of both is given below

6.2.2 UNIT TESTING

OBJECTIVE

The objective of Unit Testing is to test a unit of code (program or set of programs) using the Unit Test Specifications, after coding is completed. Since the testing will depend on the completeness and correctness of test specifications, it is important to subject these to quality and verification reviews.

INPUT

- Unit Test Specifications
- Code to be tested.

TESTING PROCESS

- Checking for availability of Code Walk-thru reports which have documented the existence of and conformance to coding standards
- Review of Unit Test Specifications
    - Verify the Unit Test Specifications conform to the program
    - Verify that all boundary and null data conditions are included

TECHNIQUES

Test Walk-through

This method of reviewing modules, a check for testability, is done by mentally executing the code with example test cases. The advantage is that the programmer can ensure that the path taken is always the one intended and that the values of variables are always sensible. This is not a proven method and can be used only by persons with knowledge of the particular language or application. It is recommended that a number of test cases from the Unit Test Specifications be utilized for Test Walk-through. Two strategies are defined for testing modules by actual execution: the top down approach and the bottom-up approach.

In the top-down approach, the main routine is tested first, with the subroutines being substituted by dummy or null routines called stubs. As testing progresses, the stubs are replaced by real routines that in turn may call further stubs representing lower-level modules in the module hierarchy.

In the bottom-up approach, testing starts at the lowermost level and then proceeds to higher-level programs or modules. The testing team directly calls the module to be tested, which may call other pre-tested modules.

6.2.3 INTEGRATION TESTING

After completion of our module along with testing, modular coding strategy was used. After integrating the module with the complete application, time was given to our team to test their part of module completely and thoroughly. As the whole application is divided into several modules, there were a lot of variable names and function names, which were common to all the modules. There existed a lot of session variables, which we had to incorporate into our module, but as different modules were being developed simultaneously, we had to hard code things in place of the session variables in our module. So, at the time of integration a lot of hard coded things had to be removed and session variables were replaced.

6.3 TEST CASES

Serial Number	Test Case Description	Expected Result	Actual result
1.	No username and password.	• The landing page should be displayed. • Patient can register and login into the system. • Doctor can login into the system. • Admin can login into the system.	Same as expected.
2.	Correct username and password.	• User is redirect to the respective dashboard. • Patient can check for heart, pneumonia and diabetes problem. • Patient can predict disease after selecting list of symptoms. • Patient can provide feedback regarding the system. • Patient can take, cancel, view details of appointment. • Patient can view list of all doctors. • Patient can browse profile and able to edit it. • Doctor can manage appointment. • Doctor can send prescription. • Doctor can view list of all patients who predicted disease which was assigned to take charge by him.	Same as expected.

		• Admin can view number of predictions, patients, doctors, feedbacks. • Admin can manage doctor details. • Admin can manage patient details. • Admin can assign doctor to take charge of the particular disease.	
3.	Incorrect username and password.	• Invalid credentials error message should be displayed.	Same as expected.
4.	Fill in field of search for doctors by patient and submit.	• List of doctors should be displayed • If the doctor is not available no results will be shown. • Browse profile of doctors.	Same as expected.
5.	Fill in field of search for appointment and submit.	• List of available appointment should be displayed. • Should be able to book respective appointment if not booked by someone else.	Same as expected.
6.	Enter all parameter value for heart and diabetes prediction.	• If all parameter values are filled up system gives the output results. • If any of the parameter value is not supplied error message should be displayed.	Same as expected.
7.	Upload x –ray for pneumonia prediction.	• If x –ray image is provided system gives the output results. • If any other image is provided, warning message should be displayed.	Same as expected.
8.	Selection of symptoms for prediction of disease.	• The outcomes will be predicted disease name along with doctor suggestion. • If the outcome is negative(no disease predicted),informational message will be shown instead of doctor suggestion.	Same as expected.
9.	Fill in field of search for patient	• List of patients should be displayed.	Sane as expected.

	by doctor and submit.		
10.	Fill in field of search for doctors by admin.	• List of available doctors should be displayed.	Same as expected.
11.	Fill in field of search for patient by admin.	• List of patient should be displayed.	Same as expected.
12.	Click on Logout button.	• Popup window should be displayed and if clicked on logout ,user should be redirect to home page else if cancel button is clicked user should remain in same page .	Same as expected.

Table 1 : Test Cases

7. PROJECT TASK AND TIME SCHEDULE

The project schedule has been designed as per requirement and constraints involved. This project is schedule to be completed in about 8 months. Study and analysis has been given more emphasis. Research and database management is to be done first and well document. Debugging and testing is to be done prior to the completion of project.

TASK	**APPROX DURATION (In weeks)**
Study and Analysis	5
Data Collection	7
Personal Collection	3
Survey	3
Social mining	5
Implementation	10
Testing	9
Documentation	8
Revisions	7
Deployment	4

Table 2: Project task and time schedule

7.1 GANTT CHART

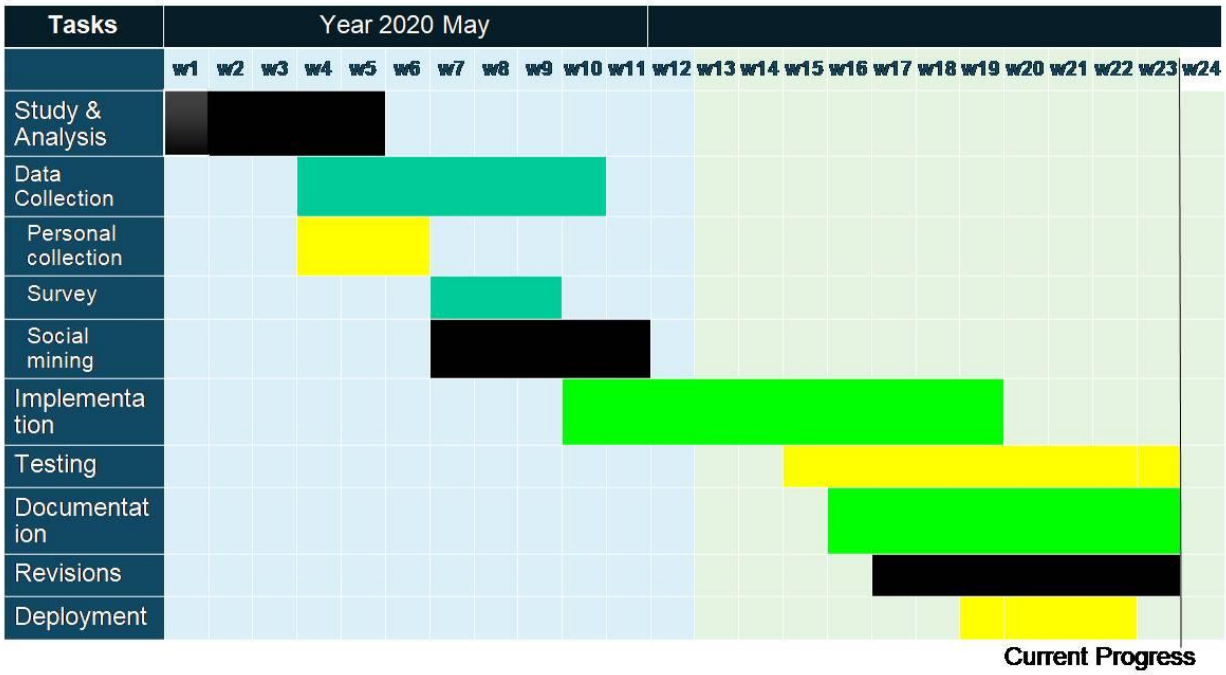

Figure 30 : Gantt chart

8. CONCLUSION

K-NN algorithm, random forest algorithm, convolution neural network and different algorithmic processes have helped to generate E-Health Care system effectively. This web-based application will be very useful because not only is it efficient but a great method to minimize the medical error and provide better clinical decision making. This system builds an environment that facilitate preliminary examination of patient health providing a functionality to check for heart, diabetes and pneumonia problem and also the examination of various disease according to symptoms provided along with doctor's suggestion to consult. The system will prove helpful to schedule an appointment for consultation between doctor and patient and also allows doctor to send online subscriptions to email as per the patient need. In today's day and age electronic health care applications are much required because of the increasing demands of new trends, technology and all of our systems are centrally controlled through ICT. This application accelerates the much-needed transformation of our health care systems and sustain access, affordability and quality for all in the near future.

9. FURTHER WORKS/FUTURE RECOMMENDATION

We can make many extensions in this project in future to make this system more reliable and trustier. The future extension of this project are as follows:

- None of payment method is integrated, integration of different payment method will enhance the system.
- Implementation of chat room for real time communication between doctor and patient.
- Making a mobile app for the user is also a first step development.
- Implementation of medical lab resources management system.

10. BIBLOGRAPHY

[1] Zhenguo Ping, Jing Chen. A personalized recommendation system based on association rules[J]. Computer integrated system, 2018:1–2.

[2] Chen L, Nayak R, Xu Y. A Recommendation Method for Online Dating Networks Based on Social Relations and Demographic Information[C]. In Proceedings of ASONAM, 2018: 407–411.

[3] Kompan M, Bieliková M. Content-Based News Recommendation [J/OL]. In Proceedings of EC-Web, 2018: 61–72.

[4] Ailin Deng, Yangyong Zhu, Bole Shi. Collaborative filtering recommendation algorithm based on item score prediction[J]. Journal of software, 2003:1621–1628.

[5] Kehan Chen, Panpan Han, Jian Wu. Recommendation algorithm for heterogeneous social networks based on user clustering[J]. Journal of Computer Science, 2019:2–5.

[6] AsadiSrinivasulu, S.Amrutha Valli, P.Hussainkhan, and P.Anitha."ASurvey on Disease Prediction in big data healthcare using extended convolutional neural network". National conference on Emerging Trends in information, management and Engineering Sciences, [2019].

[7] Nilesh Jagdish Vispute, Dinesh Kumar Sahu, Anil Rajput,"ASurvey on naive Bayes Algorithm for Diabetes Data Set Problems", International journal for research in Applied Science & Engineering Technology (IJRASET),Volume 3 issue XII,December2019.

[8] VrushaliBalpande, Rakhi Wajgi," Review on Prediction of Diabetes using Data Mining Technique", International Journal of Research and Scientific Innovation (IJRSI) |VolumeIV,IssueIA,January 2017 | ISSN2321– 2705.

[9] Rubin, J., Sanghavi, D.,Zhao,C.,Lee,K.,Qadir,A.,Xu-Wilson,M.:LargeScaleAutomated Reading of Frontal and LateralChest X-Rays Using Dual Convolutional Neural Networks (2018). arXiv preprintarXiv:1804.07839

[10] Lakhani,P.,Sundaram, B.: Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. Radiology**284** (2), 574–582 (2017)