

A Major Project Final Report on  
**DocConnect**

Submitted in Partial Fulfillment of the  
Requirements for the Degree of  
**Computer Engineering**  
under Pokhara University

Submitted by:  
**Harkesh Kathayat, 181312**  
**Samiksha Saud, 181336**  
**Bhoj Raj Joshi, 181308**

Under the supervision of  
**Dr. Ashim Khadka**

Date:  
12<sup>th</sup> July, 2023



Department of Computer Engineering  
**NEPAL COLLEGE OF  
INFORMATION TECHNOLOGY**

---

Balkumari, Lalitpur, Nepal.

## Acknowledgement

The satisfaction that accompanies the successful completion of any task would be incomplete without the mention of people whose ceaseless cooperation made it possible, whose constant guidance and encouragement crown all efforts with success. It is an utmost privilege to express our sincere regards to the project supervisor, **Dr. Ashim Khadka** for his valuable guidance, encouragement and support, throughout the duration of the project. We deeply express our sincere thanks to him for encouraging and allowing us to present the project on the topic “**DocConnect**”.

We take this opportunity to thank our Project Head, **Dr. Roshan Chitrakar** for providing us the opportunity to embark this project. We are also highly obliged to Hod, **Er. Resha Deo** and all the lecturers, who have directly or indirectly helped us to gain the successful completion of the project. We also express our respect and love to our guardians and all other family members for their love and encouragement throughout our project development period. Last but not the least, we express our thanks to all our beloved friends for their cooperation and support.

## **Abstract**

It might have occurred so many times when you or someone you know requires quick medical attention, but they are unavailable for any reason. DocConnect is a project for end-user assistance and online consultation. DocConnect is an online platform designed for the prediction of diseases and facilitating doctor consultations. Its primary purpose is to serve as a medium for patients and doctors to engage in the consultation process. DocConnect is basically a web app that enables patients to create accounts and log in to access disease information. By providing a list of symptoms, the system utilizes symptom-based algorithms to predict potential diseases. Additionally, it suggests relevant doctors specializing in the identified diseases. Doctors can also register accounts, granting them access to view patients' consultation histories. The admin oversees the management of all registered doctors and patients within the system. Both patients and doctors have the autonomy to manage their respective profiles. Ultimately, DocConnect delivers a comprehensive website capable of disease prediction, doctor suggestions, and consultation process management.

**Keywords:** DocConnect, Prediction, Consultation, Disease.

## Table of Contents

Acknowledgement .....	ii
Abstract .....	iii
List of Table.....	v
List of Figures .....	vi
1. Introduction .....	1
1.1 Problem Statement.....	1
1.2 Project Objectives .....	2
1.3 Significance of study .....	2
2. Literature Review.....	3
3. Methodology.....	5
3.1 Software Development Lifecycle .....	5
3.2 System Design and UML diagrams.....	7
3.2.1 Use case Diagram.....	7
3.2.2 Entity Relation Diagram .....	8
3.2.3 Activity diagram.....	9
3.2.4 Architecture of the system .....	11
3.3 Machine learning Algorithm .....	12
3.3.1 Gaussian Naïve Bayes Algorithm .....	12
3.3.2 Disease Prediction .....	12
3.3.3 Dataset .....	14
3.4 Tools and Technologies to be used.....	15
3.4.1 Tools .....	15
3.4.2 Technology .....	15
4. Conclusion.....	17
5. Future work / Recommendations .....	18
6. Feasibility Analysis.....	19
6.1 Task and Time schedule.....	19
6.2 Work Break Down.....	20
7. Bibliography/References .....	21
Appendix .....	22

**List of Table**

Table 1 Tools ..... 15

Table 2 Task and Time schedule ..... 19

Table 3 work division ..... 20

## List of Figures

Figure 1 Incremental Process .....	6
Figure 2 Use case diagram.....	7
Figure 3Entity relation diagram .....	8
Figure 4 Activity diagram for patient .....	9
Figure 5 Activity diagram for Doctor .....	10
Figure 6 Overall system architecture .....	11
Figure 7 Machine Learning Block diagram .....	13
Figure 8 Dataset .....	14
Figure 9 Dashboard .....	22
Figure 10 Consultant process .....	22
Figure 11 Disease prediction.....	23
Figure 12 Confusion matrix .....	24

# **1. Introduction**

This project deals with online platform for Disease prediction and consultant process. In this section, problem statement, objectives and significance of study will be described in detail.

## **1.1 Problem Statement**

The main problem that the DocConnect aims to address is the inefficiency and inconvenience of the traditional appointment scheduling process for healthcare providers and patients.

The other problems regarding this system are explained as:

- The whole process of booking an appointment is quite a tedious task. Going to the hospital, standing in queue and booking the appointment takes up half of the time of your visit.
- Traditional medical diagnosis methods based on manual analysis of symptoms can be time-consuming and error-prone.
- The medical history of a patient is not known to the doctor until they meet the patient and read it from the hospital book on the spot.
- Identifying the best doctor for a specific disease can be challenging for patients, especially those with limited access to information about doctors.

## **1.2 Project Objectives**

The main purpose of DocConnect is to simplify the diseases prediction and consultation process for patients and healthcare providers.

Furthermore, the DocConnect project can be used for the following objectives.

- To develop a user-friendly interface for patients to input their symptoms and receive possible predictions of their diseases.
- To develop an effective and efficient system for suggesting doctors based on speciality.
- To provide consultation setup for both doctor and patients.
- To provide patients with information about suggested doctors, including qualifications, experience, and reviews, to help them make informed decisions about their healthcare.
- Establish a mechanism for gathering user feedback, implementing necessary enhancements, and continuously improving the functionality and features of the platform.

## **1.3 Significance of study**

The study of the DocConnect project is significant as it enhances healthcare accessibility by providing easy access to disease information and remote doctor consultations. It enables early disease detection, leading to timely interventions and improved treatment outcomes. The project optimizes resource allocation by suggesting specialized doctors based on predicted diseases. It promotes efficient communication between doctors and patients, improving engagement and decision-making. The collected consultation data offers valuable insights for research and analysis, contributing to medical knowledge and evidence-based practices. Through continuous user feedback and improvements, the project advances healthcare services, benefiting patients and providers alike, and ultimately improving healthcare delivery. [1]



## 2. Literature Review

To create this project, we looked at few publications from various external sources. We examined the relationship between various algorithms' performance in various disease prediction scenarios. So, in this section we will be reviewing some of such existing systems that resemble this project.

Anand kumar, Prakash U M, Ganesh Kumar Sharma published a paper [2]. This paper looks at the use of Machine Learning to develop a disease prediction and doctor recommendation system. Different classification algorithms like Logistic Regression, Random Forest Classifier, KNN and Naïve Bayes are used to predict a person's disease based on their symptoms and then recommends which type of doctor to consult. This system is used by end-users. An interactive interface is built as front-end and is connected to the Server. This system might have a significant impact on how doctors treat patients in the future. But, due to the complexity and variety of diseases, there may be possible accuracy issues as well as bias in the data used to train the algorithm.

Gomathy, C K. (2021) [3]. In this study, we discuss the application of machine learning to predict diseases from patient symptoms. It determines the probability of what disease could be present by using the supervised machine learning algorithm like Naive Bayes classifier. Accurate analysis can help in early identification and improve patient care along with the development of biomedical and healthcare data. It also shows how the linear regression and decision tree algorithms can be used to predict specific diseases like Diabetes, Malaria, Jaundice Dengue or Tuberculosis. The advantages of using Machine Learning model includes the ability to accurately analyse medical data, leading to early detection and better patient care. Additionally, it can be used to predict specific diseases with a high degree of accuracy. However, this approach needs a lot of data for the algorithms to function well, and bias might result if insufficient varied datasets are used.

"Medical Disease Using Machine Learning Algorithms" [4]. A standard method of diagnosis involves a patient seeing a doctor, going through several tests, and then coming to a decision. It takes a long time to do this task. This project suggests an automated disease prediction system to reduce the time needed for the initial disease prediction process, which depends on user input. It is designed in a way such that when the user is introduced to the chatbot system, they are given the choice of receiving an estimation or prediction of their disease based on the data they have

provided to the chatbot. Data is gathered from Columbia University to help with disease prediction, and Kaggle provides a source for the diseases' related symptoms. There are several machine learning algorithms used, such Naive Bayes, Random Forest Classifier, K-Nearest Neighbours, and Support Vector Machine classifier. However, when fewer symptoms are submitted, the accuracy will be reduced.

### **3. Methodology**

#### **3.1 Software Development Lifecycle**

For the development of this project, we are using Incremental model. In this model the requirements are broken down into different standalone modules which work as individual increments. Each increment goes through the requirement analysis, design, coding and testing phases.

- **Requirement analysis:**

In this phase, requirements are identified by a thorough analysis of the project. A SRS (System Requirements specification) document is generated as output of this phase.

- **Design:**

Initially, the requirements identified and documented in the previous stages are studied and system design is carried out according to it.

- **Coding:**

In this phase, coding is done based on the system design developed. A working product is developed by the completion of this phase.

- **Testing:**

The completed system is tested and any required change is implemented to create the finished product that is ready for deployment.

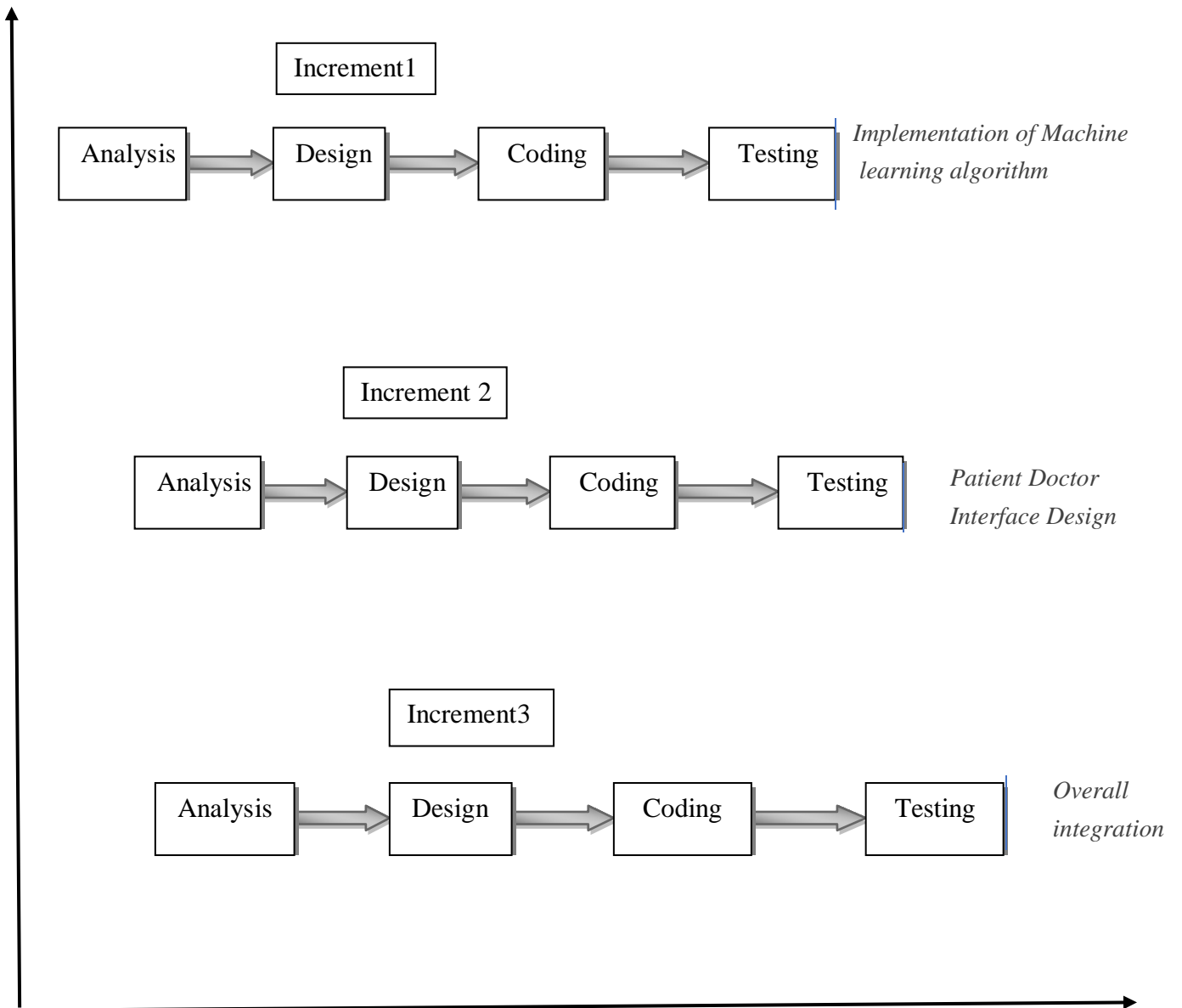


Figure 1 Incremental Process

## 3.2 System Design and UML diagrams

### 3.2.1 Use case Diagram

Use-cases is used to explain sequence of action performed in a system. This diagram identify the interactions between the system and its actors.

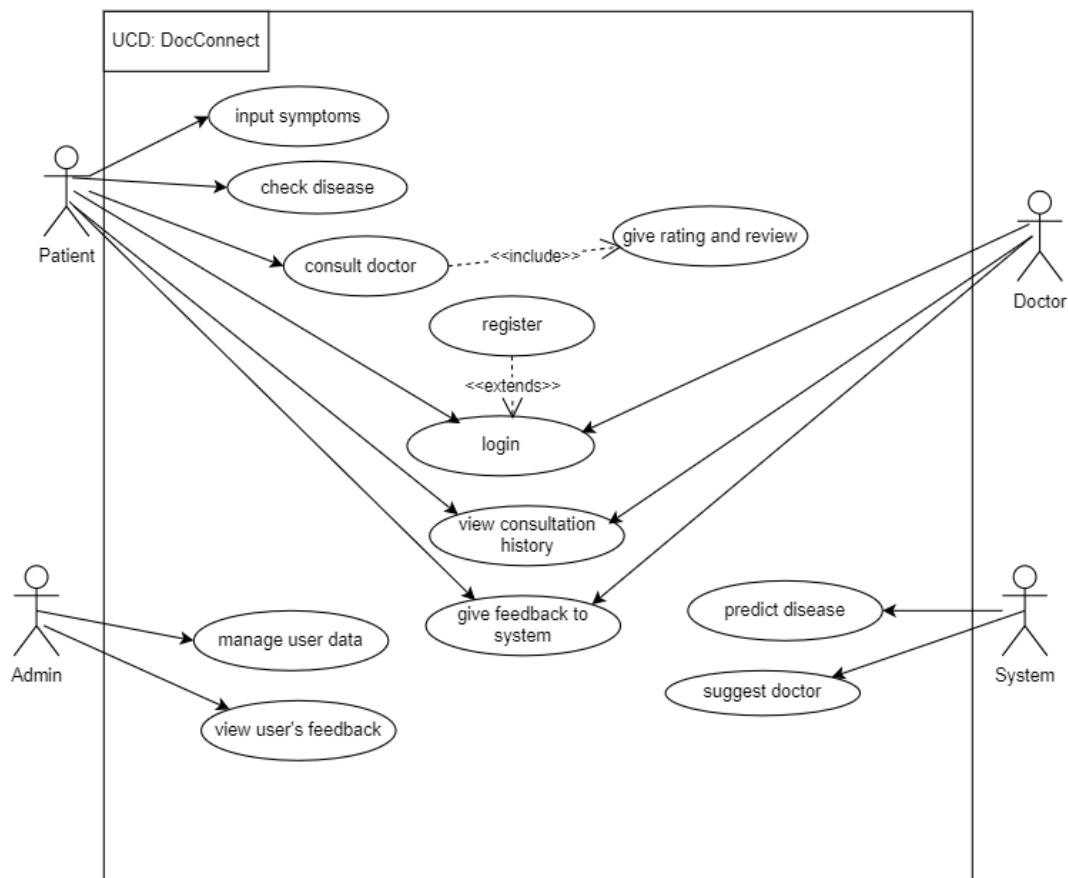


Figure 2 Use case diagram

### 3.2.2 Entity Relation Diagram

Entity Relationship Diagram is used to for graphical representation of relationship between attributes. It illustrates how “entities” such as patient, doctor or concepts relate to each other within a system.

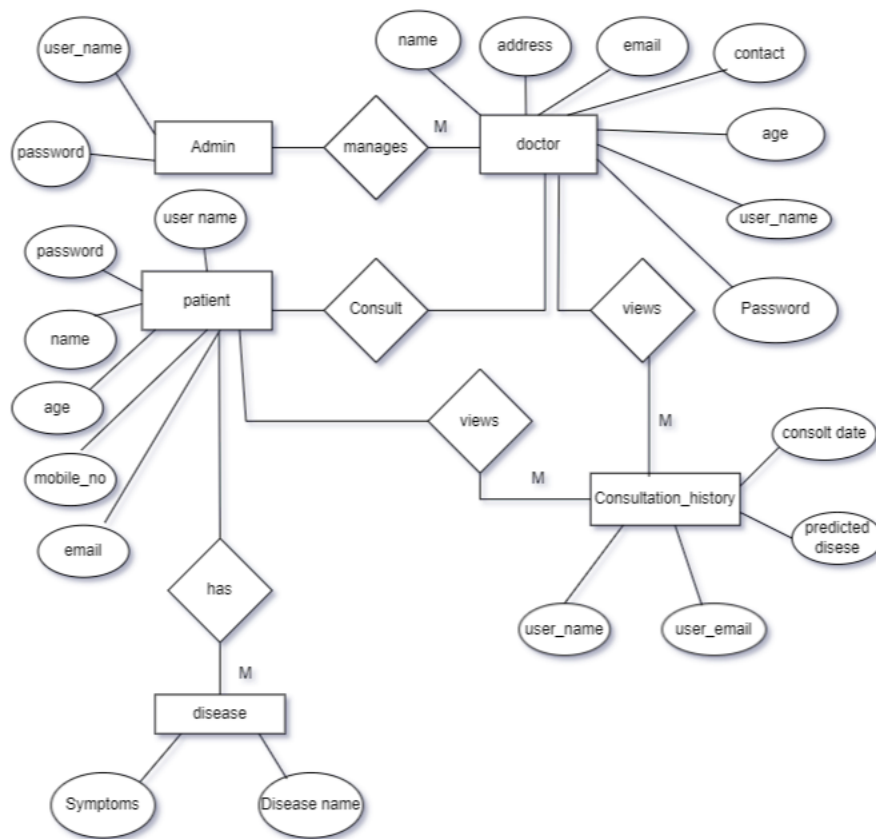


Figure 3Entity relation diagram

### 3.2.3 Activity diagram

Activity diagrams map out process flows in a way that's easy to understand. It provides a view of the behavior of a system by describing the sequence of actions in a process.

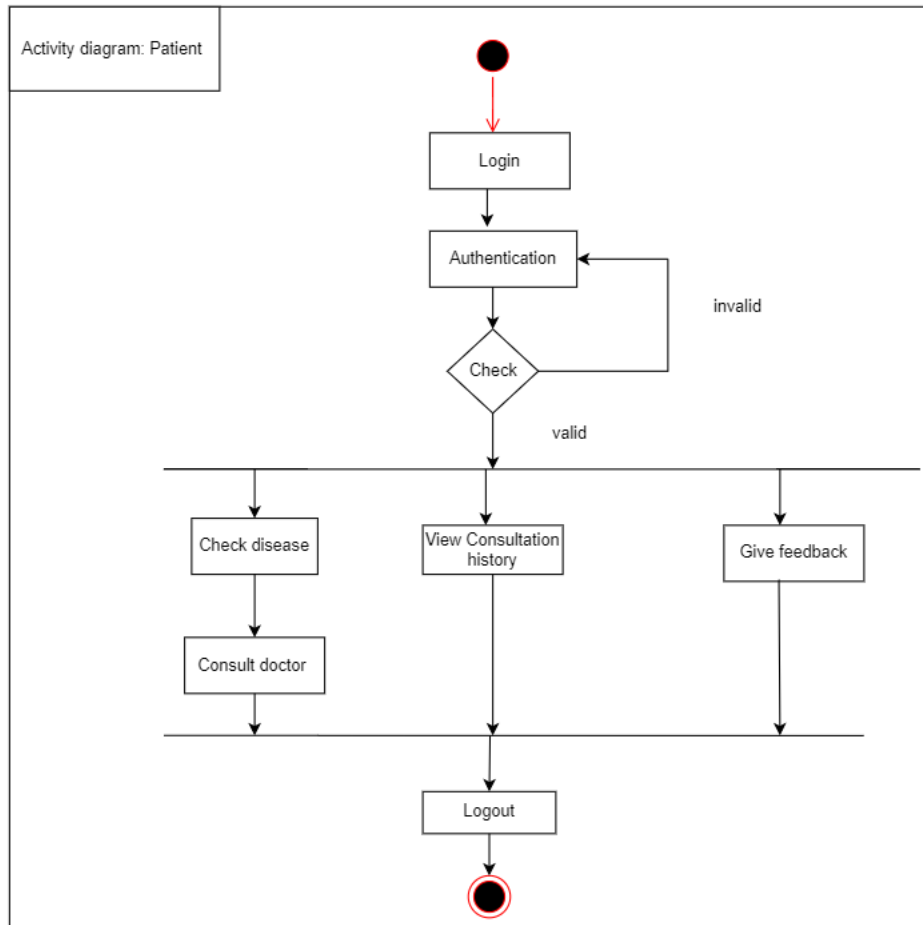


Figure 4 Activity diagram for patient

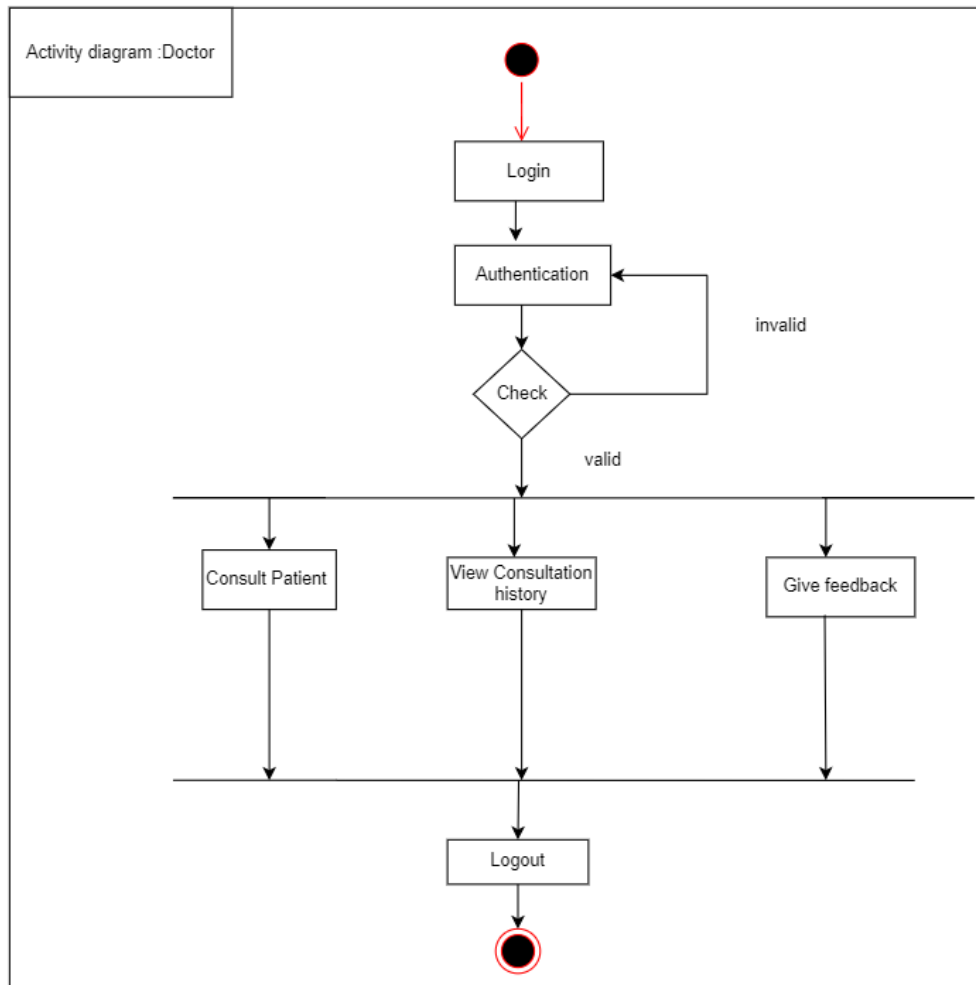


Figure 5 Activity diagram for Doctor



### 3.2.4 Architecture of the system

The architecture visually depicts how data moves through a system, illustrating the key steps in a clear and organized manner. It presents a systematic and sequential representation of the data flow within the system.

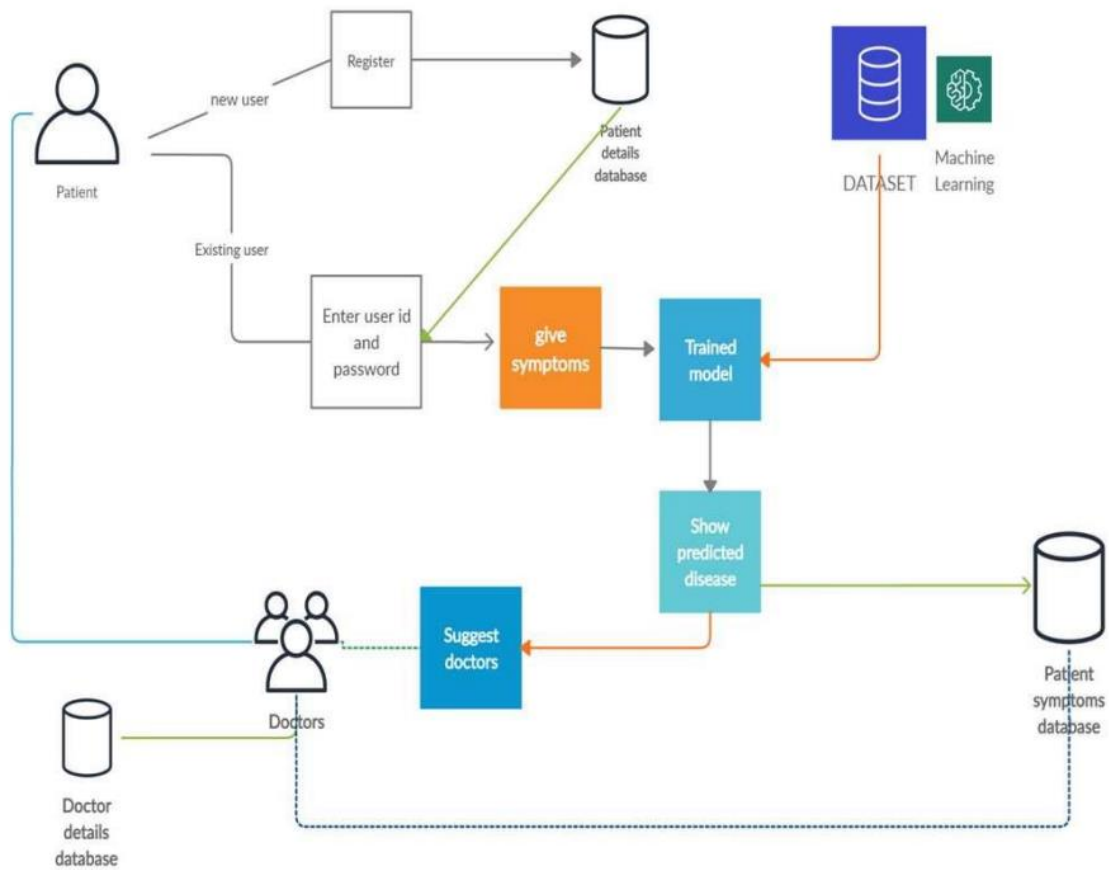


Figure 6 Overall system architecture

### **3.3 Machine learning Algorithm**

A prediction system based on machine learning utilizes advanced algorithms and models to make predictions or forecasts based on input data. These systems are trained on historical data to identify patterns and relationships, allowing them to generate predictions or estimates for future or unknown data points.

#### **3.3.1 Gaussian Naïve Bayes Algorithm**

Gaussian Naive Bayes is a probabilistic classification algorithm that uses the Bayes theorem with strong independence assumptions. It is a variant of Naive Bayes that supports continuous data and corresponds to the Gaussian normal distribution. Each parameter (also known as a feature or a predictor) is assumed to have an independent ability to predict the output variable by Gaussian Naive Bayes. It is a fast and flexible model that produces highly reliable results on large data sets. There is no need to spend a lot of time towards training. It also improves grading performance by removing unimportant specifications. It is a probabilistic machine learning algorithm that internally uses Bayes Theorem to classify the data points.

Mathematically, the conditional probability of A given B can be denoted as

$$P(A|B) = P(B|A) P(A) / P(B)$$

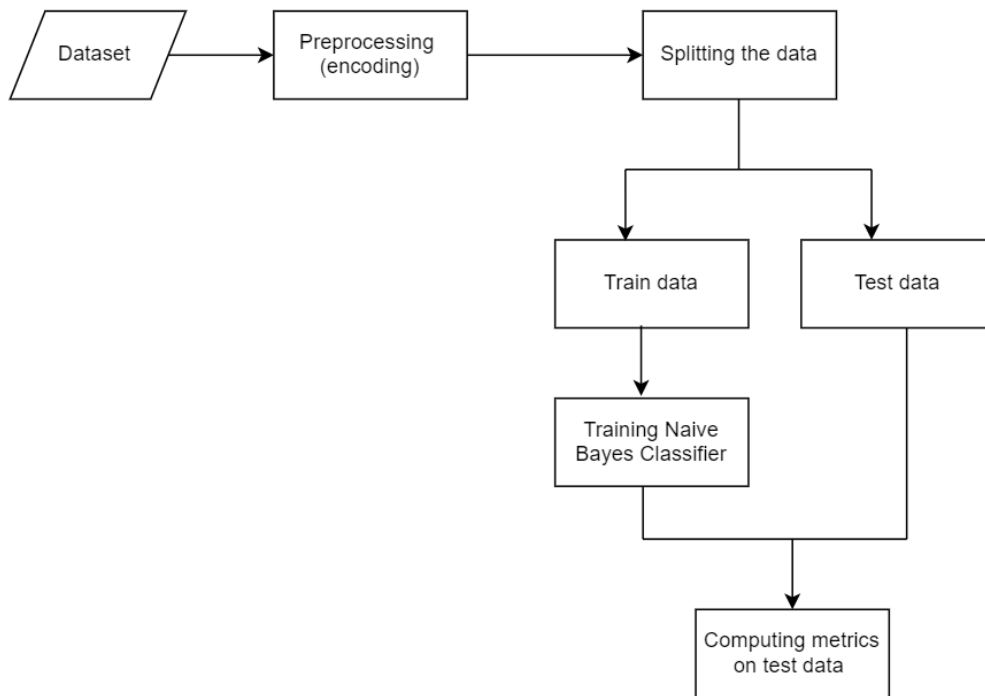
#### **3.3.2 Disease Prediction**

Gathering the Data:

Data preparation is the primary step for any machine learning problem. We used a dataset from Kaggle for this problem. This dataset consists of various diseases and their respective symptoms.

Cleaning the Data:

Cleaning is the most important step in a machine learning project. The quality of our data determines the quality of our machine-learning model. So, it is always necessary to clean the data before feeding it to the model for training. In our dataset all the columns are string type and are encoded to numerical form using an encoder.



*Figure 7 Machine Learning Block diagram*

#### Model Building:

After gathering and cleaning the data, the data is ready and can be used to train a machine learning model. We used this cleaned data to train the Gaussian Naive Bayes Classifier.

#### Inference:

After training the models we will be predicting the disease for the input symptoms. This makes our overall prediction more robust and accurate. [5]

### 3.3.3 Dataset

Data collection has been done from the Kaggle to identify the disease. Kaggle is an online community of data scientists and machine learning practitioners who participate in machine learning competitions, collaborate on data science projects, and share knowledge and expertise on data science topics. [6]

Here the real symptoms of the disease are collected i.e. no dummy values are entered.

This csv file contains 132 types of different symptoms and their corresponding disease (40 class of general disease).

	Disease	Symptom_1	Symptom_2	Symptom_3	Symptom_4	Symptom_5	Symptom_6	Symptom_7	Symptom_8	Symptom_9
0	Fungal infection	itching	skin_rash	nodal_skin_eruptions	dischromic_patches	NaN	NaN	NaN	NaN	NaN
1	Fungal infection	skin_rash	nodal_skin_eruptions	dischromic_patches	NaN	NaN	NaN	NaN	NaN	NaN
2	Fungal infection	itching	nodal_skin_eruptions	dischromic_patches	NaN	NaN	NaN	NaN	NaN	NaN
3	Fungal infection	itching	skin_rash	dischromic_patches	NaN	NaN	NaN	NaN	NaN	NaN
4	Fungal infection	itching	skin_rash	nodal_skin_eruptions	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...	...	...
402	Impetigo	high_fever	blister	red_sore_around_nose	yellow_crust_ooze	NaN	NaN	NaN	NaN	NaN
403	Impetigo	skin_rash	blister	red_sore_around_nose	yellow_crust_ooze	NaN	NaN	NaN	NaN	NaN

Figure 8 Dataset

## 3.4 Tools and Technologies to be used

### 3.4.1 Tools

*Table 1 Tools*

Tools	Purpose
Jupyter Notebook	IDE for model training
VS Code	IDE for writing code
GitHub	Platform for a team of developers to work concurrently on same project
Draw.io	Tool for diagrams

### 3.4.2 Technology

#### Python

Python is an interpreted, high-level, general-purpose programming language. Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural. It also has a comprehensive standard library.

#### Django

Django will be used for backend process. Django is a high-level web framework for building web applications in Python. It was created to help developers take a pragmatic approach to web development and encourage the use of clean, reusable code.

#### Pandas

Pandas is an open-source data analysis and manipulation library for the Python programming language. It provides data structures for efficiently storing and manipulating large datasets, and tools for working with structured data such as tabular data, time series data, and more.

#### Scikit-learn

Scikit-learn (Sklearn) is Python's most useful and stable machine learning library. It offers a range of powerful machine learning and statistical modeling methods, including classification, regression, clustering and reduction of dimensionality via a Python consistency interface. This library is based on NumPy, SciPy and Matplotlib, and is largely written in Python.

## **PostgreSQL**

PostgreSQL is a powerful and open-source relational database management system (RDBMS) that is known for its robustness, scalability, and extensive features. It is designed to store and manage large amounts of structured data efficiently. PostgreSQL supports a wide range of SQL features and provides additional capabilities like transactional integrity, concurrency control, and data integrity.

### **Front-end**

- ✓ HTML
- ✓ CSS
- ✓ JavaScript
- ✓ Bootstrap

## **4. Conclusion**

In conclusion, the DocConnect project is a significant development in the healthcare sector that aims to enhance accessibility to healthcare services, improve disease detection, and streamline the doctor-patient consultation process. By providing an online platform for disease prediction, doctor suggestion, and consultation management, DocConnect offers convenience and efficiency to both patients and doctors. The project leverages symptom-based algorithms and advanced technologies to predict diseases based on patient-provided symptoms, connecting them with relevant specialists for further consultation.

## **5. Future work / Recommendations**

Although we try to cover main processes, there are several areas for further work and recommendations for the future development of this project:

- Expansion of disease prediction capabilities,
- Chat system for continuous interaction of patient with doctor,
- Enhanced security and privacy measures,
- Collaboration with healthcare institutions and organizations.

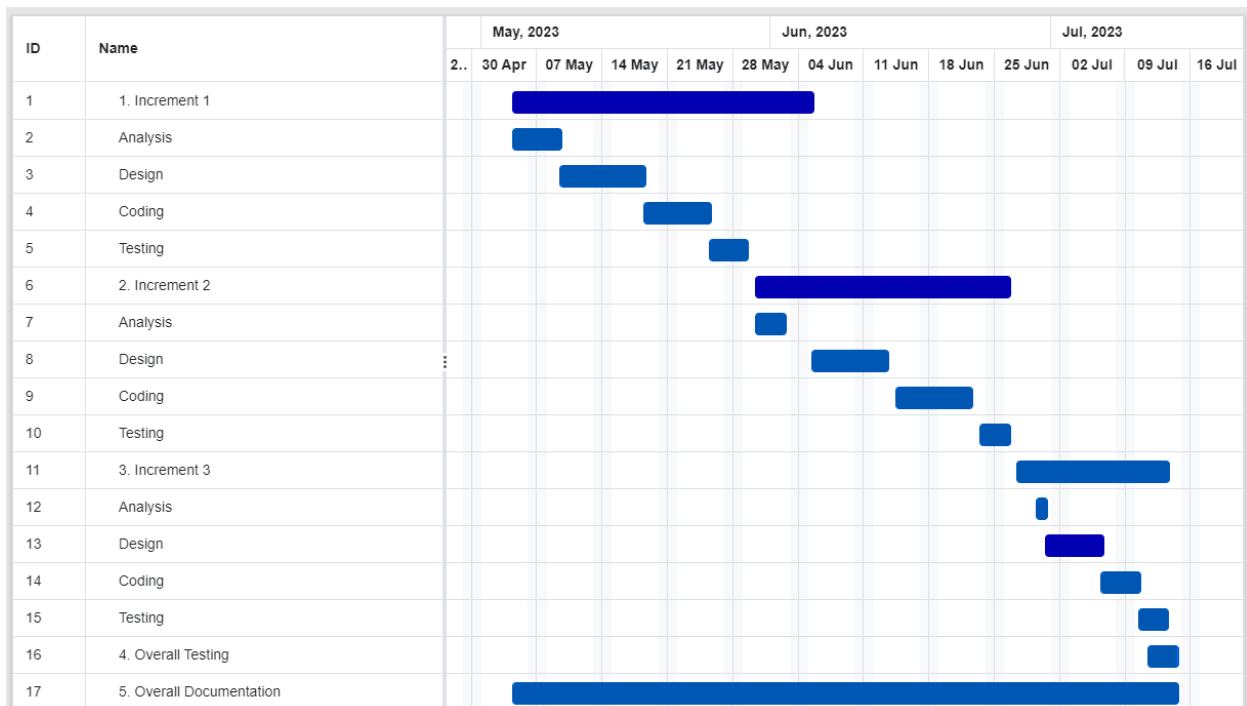


## 6. Feasibility Analysis

### 6.1 Task and Time schedule

The time schedule for the project had been estimated according to the stages involved in the development of the system. Our project was estimated to be completed in around 3 months. We have completed all three increments with in estimated time.

Table 2 Task and Time schedule



## 6.2 Work Break Down

We have divided our project mainly into three parts and assigned to each team member as in table below:

*Table 3 work division*

<b>S.N.</b>	<b>Task</b>	<b>Member</b>
<b>1.</b>	Machine Learning	Harkesh Kathayat, Samiksha Saud
<b>2.</b>	Backend (Django integration)	Bhoj Raj Joshi, Harkesh Kathayat
<b>3.</b>	Frontend	Samiksha Saud, Bhoj Raj Joshi

## 7. Bibliography/References

- [1] ccioliveadm2017, "Continuous care for health." <https://www.continuouscare.io/blog/benefits-of-online-appointment-scheduling-software-for-doctors-clinics/>. (Accessed 7<sup>th</sup> May, 2023.)
- [2] Anand kumar, Prakash U M, Ganesh Kumar Sharma, "Disease Prediction and Doctor Recommendation System using Machine Learning Approaches".  
International Journal for Research in Applied Science and Engineering Technology.  
9.10.22214/ijraset.2021.36234. (Accessed 7th May, 2023.)
- [3] Gomathy, C K. (2021). "The Prediction of Disease using Machine Learning". (Accessed 4th June, 2023)
- [4] Kunal Takke, Rameez Bhaijee, Avanish Singh; "Medical Disease Using Machine Learning Algorithms"; 2022 International Journal for Research in Applied Science & Engineering Technology  
(IJRASET), May 2022. (Accessed 15th June, 2023)
- [5] venugopalkadamba, "Disease Prediction Using Machine Learning" GeeksforGeeks. (Accessed 11<sup>th</sup> May, 2023)
- [6] NIYAR R BARMA, Faizal Karim, Krish Sharma. "Symptom2Disease, Diseases and Natural Language Symptom Descriptions" [www.kaggle.com/datasets/](https://www.kaggle.com/datasets/). (Accessed 11<sup>th</sup> May, 2023)

## Appendix

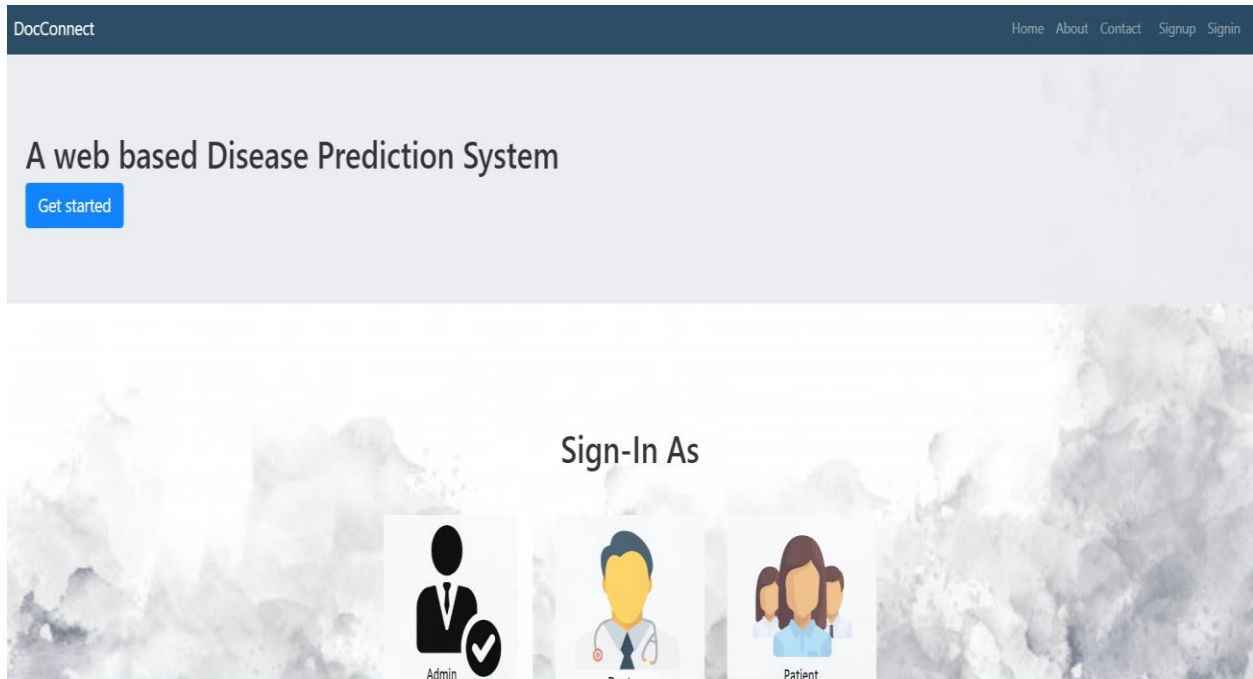


Figure 9 Dashboard

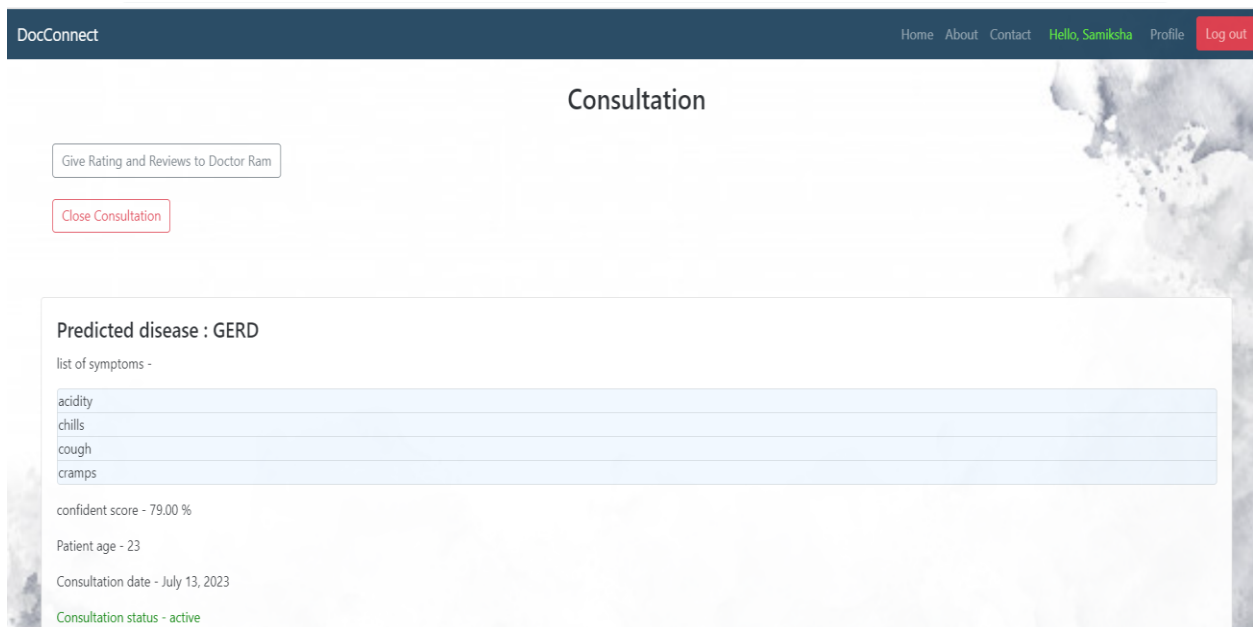


Figure 10 Consultant process

Symptoms list -

acidity

chills

cough

cramps

Predict

Patient name : Samiksha      Age : 23

predicted disease is : GERD

confidence score of : 

79%

[Click here to know more about GERD](#)

Figure 11 Disease prediction

