Birla Institute of Technology & Science, Pilani
Hyderabad Campus

Advanced Database Systems - CSG516

Lab Assignment 2

Design and develop a Health Recommendation System

Group number - 4

Name	Campus ID	Email id
Srinidhi P Katte	2022H1030075H	h20221030075@hyderabad.bits-pilani.ac.in
Shivam Rajesh Rajput	2022H1030058H	h20221030058@hyderabad.bits-pilani.ac.in
Shashank S	2022H1030067H	h20221030067@hyderabad.bits-pilani.ac.in
Ajinkya Medhekar	2022H1030099H	h20221030099@hyderabad.bits-pilani.ac.in

Introduction

A recommendation system is a type of information filtering system that predicts and suggests items that a user may be interested in based on their previous interactions and behavior. It analyzes the user's preferences and makes personalized recommendations by using machine learning algorithms and data mining techniques. The aim of a recommendation system is to provide relevant and useful suggestions to users, thereby enhancing their experience and increasing the likelihood of them engaging with the recommended items. A health recommendation system is a type of recommendation system that provides health-related recommendations to users or medical staff based on the various health parameters, medical history, lifestyle, and other relevant factors. These systems may use various data sources such as electronic health records, wearable devices, and user-reported data to generate recommendations for disease prevention, treatment, and management. The ultimate goal of a health recommendation system is to improve the overall health and well-being of individuals by providing them with tailored, evidence-based recommendations that are aligned with their unique needs and preferences.

We have developed a health recommendation system, which can be used by doctors and other medical staff in hospitals, to know about disease predictions of the patients in the hospital and it will enable them to decide treatment plans for the patients and stop further spreading of diseases. We have used open source data available on the Internet to train our Machine learning (ML) models and disease prediction for new patients will be based on the output from the ML models. Our dataset consists of information of various vital health parameters like blood pressure, cholesterol, glucose level etc and also we have used blood smear specimen image, CT scan image and activity graph of the patients.Our application is deployed in **AWS Cloud**. We have used **AWS S3** for storing images and CSV files related to patients and **AWS DynamoDB** for storing other patient related information and URL of files stored in S3. We have trained our ML models using **scikit learn** and **tensor flow** libraries which will give predictions of diabetes, heart disease, covid, pneumonia, malaria and mental depression based on the patient information. We have used **Python Flask** library to develop our web application for displaying health

recommendation information. Code for our project can be found in https://github.com/srinidhikatte/health-recommendation-system.

Demo of our application can be found in https://adbmsla2.s3.ap-south-1.amazonaws.com/adbms_la2_group4_demo.mkv

Application architecture

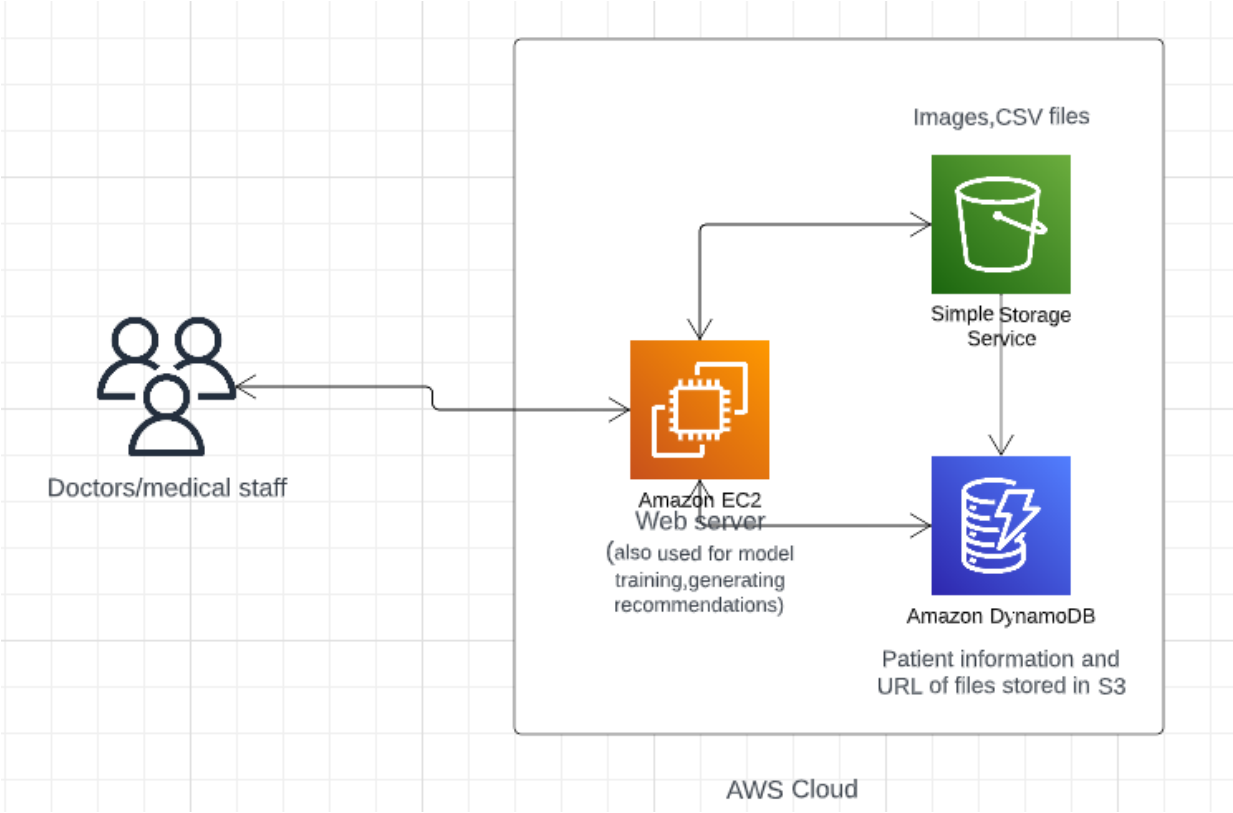


Fig 1: Application architecture

We have used below datasets

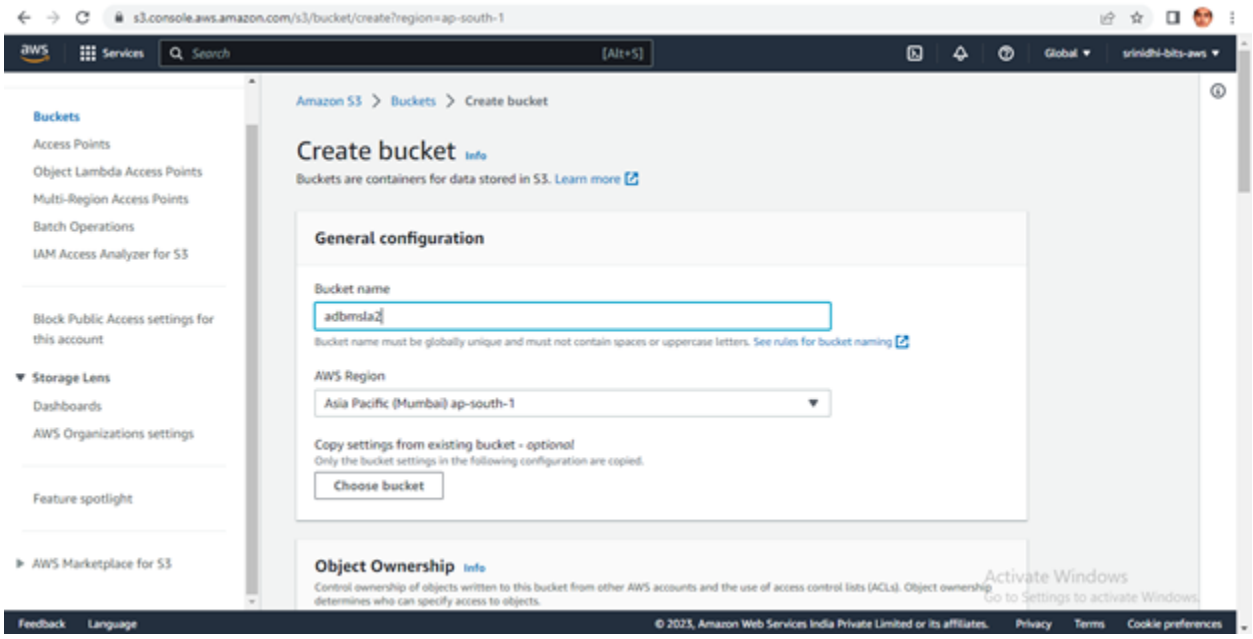
Disease	**Dataset type**	**Source**
Heart disease	CSV	https://archive.ics.uci.edu/ml/datasets/Heart+Disease
Diabetes	CSV	https://www.kaggle.com/datasets/mathchi/diabetes-data-set?resource=download
Covid and Pneumonia	Image	https://www.kaggle.com/datasets/pranavraikokte/covid19-image-dataset
Malaria	Image	https://lhncbc.nlm.nih.gov/LHC-downloads/downloads.html#malaria-datasets

Mental health	CSV(actig raph)	https://datasets.simula.no/depresjon/#dataset-details

We have used AWS S3 to store blood smear specimen image (malaria), CT scan image (covid and pneumonia) and activity graph (mental health) of the patients.

Steps for Creating AWS S3 bucket

1. Login to AWS management console and navigate to S3
2. Click on “Create Bucket” option , give name for our S3 bucket which should be globally unique , select nearest AWS region



3. We keep other bucket settings as default and create our S3 bucket.

Why have we used S3 to store images and CSV files?

1. AWS S3 is a highly scalable, cost effective and reliable object storage service for storage and retrieval of huge amount of data
2. AWS S3 may be a better option than HDFS due to its design and features that are optimized for storing large numbers of small files (images and CSV files) while HDFS is a distributed file system that is optimized for large files, and is not designed to handle

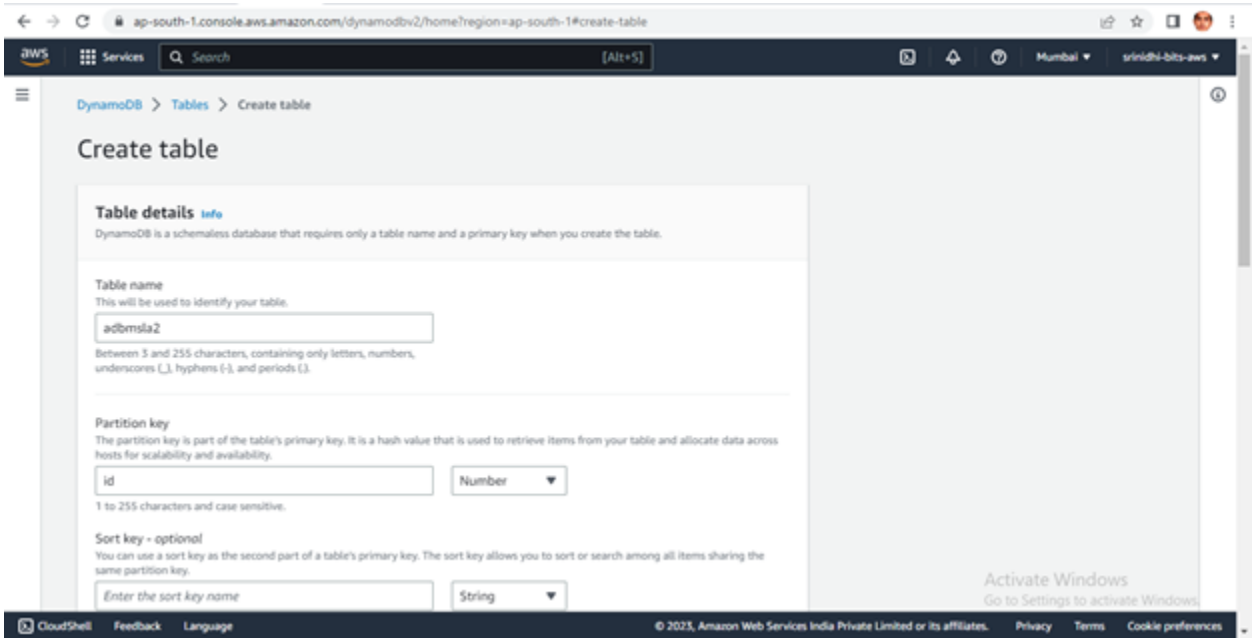
small files efficiently. This is because HDFS has a high overhead for each file, including replication, namespace management, and block management, which can cause performance issues when storing and accessing a large number of small files.

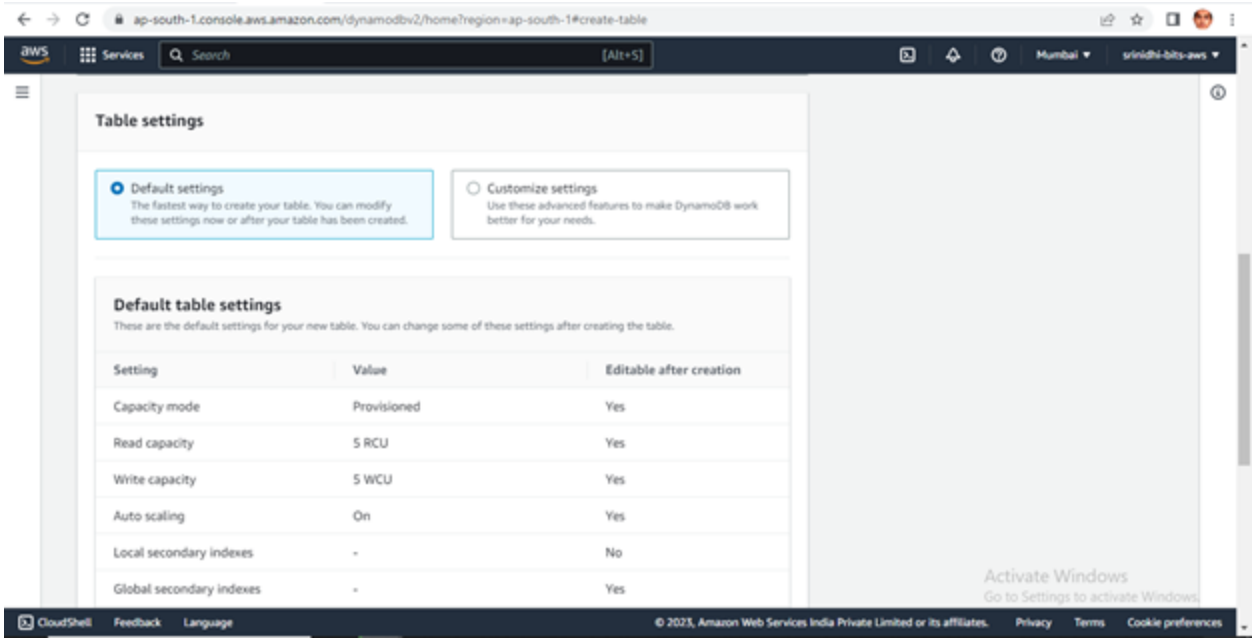
3. AWS S3 requires no installation and less maintenance overhead when compared to HDFS.

We have used AWS DynamoDB to store patient attributes like age,gender,blood pressure, glucose level, insulin, cholesterol etc and S3 URL of corresponding blood smear specimen image (malaria), CT scan image (covid and pneumonia) and activity graph (mental health) .

Steps for creating DynamoDB table

1. Login to AWS management console and navigate to DynamoDB service
2. Provide name to DynamoDB table , mention partition key as Patient ID and other settings kept to default





Why have we used DynamoDB for storing patient information?

1. DynamoDB is a NoSQL database. In our health recommendation application, the schema of a table is not fixed, since in a hospital , different patients might undergo different medical tests and we might need to store different parameters for different patients, fixing a strict schema for the patient table is not possible, hence NoSQL database like Dynamo DB is more suitable.
2. DynamoDB is useful when dealing with large volumes of semi structured, and unstructured data and suitable for applications requiring high availability and low latency.

We have used an AWS EC2 instance as a web server and we have used Python flask library for creating our web application. Same EC2 instance is also used for model training, generating datasets for testing and their prediction results.

Application design and development

Our health recommendation system application has below modules

1. Machine learning model training
2. Generating random patient data for testing and getting prediction results
3. Web app for displaying health recommendations

1.Machine learning model training

As mentioned in the previous section, we have used AWS S3 for storing images and CSV files and AWS Dynamo DB for storing patient information and URL of files stored in S3. We have collected data related to various diseases like heart disease, diabetes, malaria, covid and mental depression and stored the data in S3 and Dynamo DB. We have then used scikit learn and tensor flow libraries to train classification ML models using our dataset, which will be able to predict about the presence or absence of the disease provided all necessary patient parameters.

Our code can be found in below link

https://github.com/srinidhikatte/health-recommendation-system/tree/main/training

Trained model files were stored on EC2 instance for further testing and generating prediction results.

2.**Generating random patient data for testing and getting prediction results**

We generated patient records randomly by considering statistical features of attributes like age,blood pressure, glucose, cholesterol etc and random blood smear specimen image (malaria), CT scan image (covid and pneumonia) and activity graph (mental health) was chosen. Images and CSV files were stored in S3 bucket while patient attributes and file's S3 URL was stored in DynamoDB table.

We gave the randomly generated patient records as the input to the trained models and the corresponding results were appended to patient records in DynamoDB table.

Corresponding code can be found in below link

https://github.com/srinidhikatte/health-recommendation-system/tree/main/generate_data_test

3.Web app for displaying health recommendations

We have developed a web application using Python flask. Front end was developed using HTML and CSS. Our web application consists of below modules

3.1 Login page - For doctors and medical staff to login to our application

3.2 Analysis - Provides graphical analysis of the number of patients potential to suffer from various diseases. Provides visualizations of the number of patients potential to suffer from various diseases based on age and gender. Visualizations were developed using Python plotly library.

3.3 Patient information - Provides all the information of patients potential to suffer from various diseases.

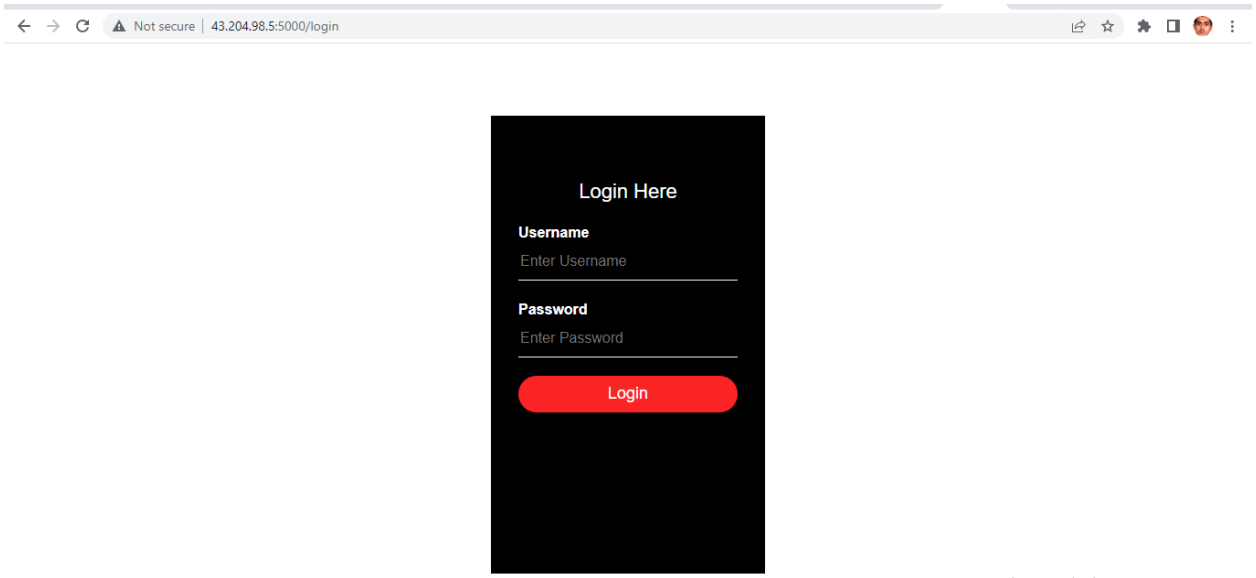
3.4 Patient diagnosis- Provided information about different patient health parameters, this module will use trained ML models to predict diseases for which the patient is susceptible to.

Our web application code can be found in below link

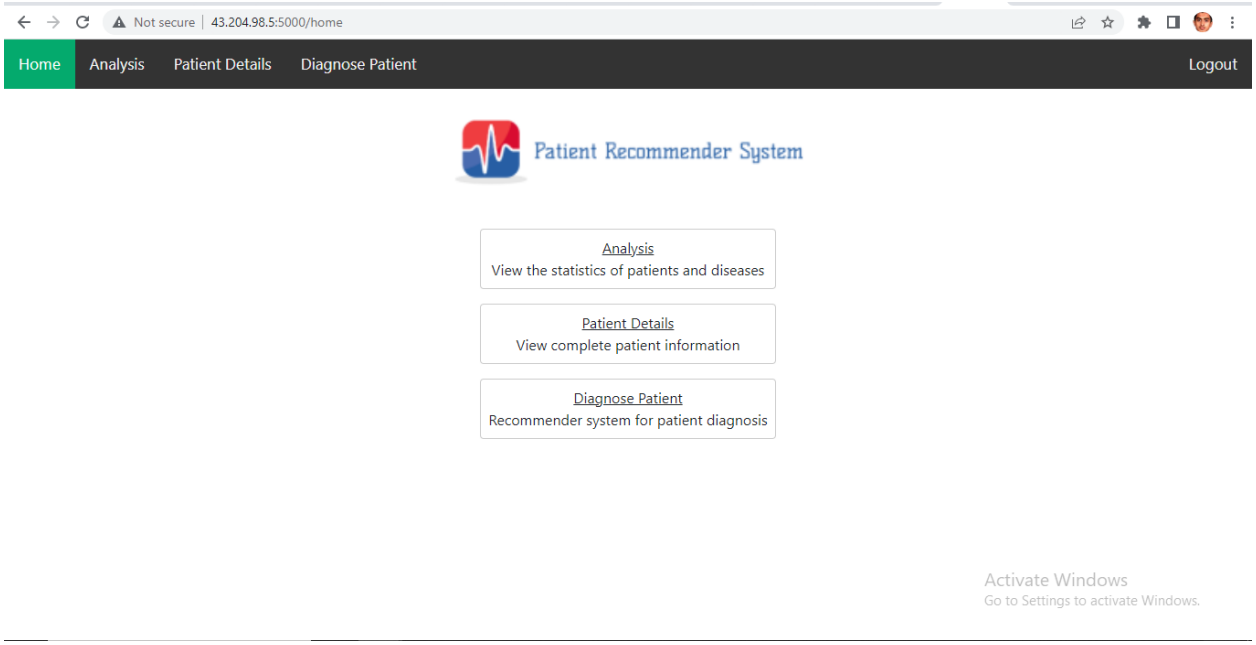
https://github.com/srinidhikatte/health-recommendation-system/tree/main/health_app

Application screenshots

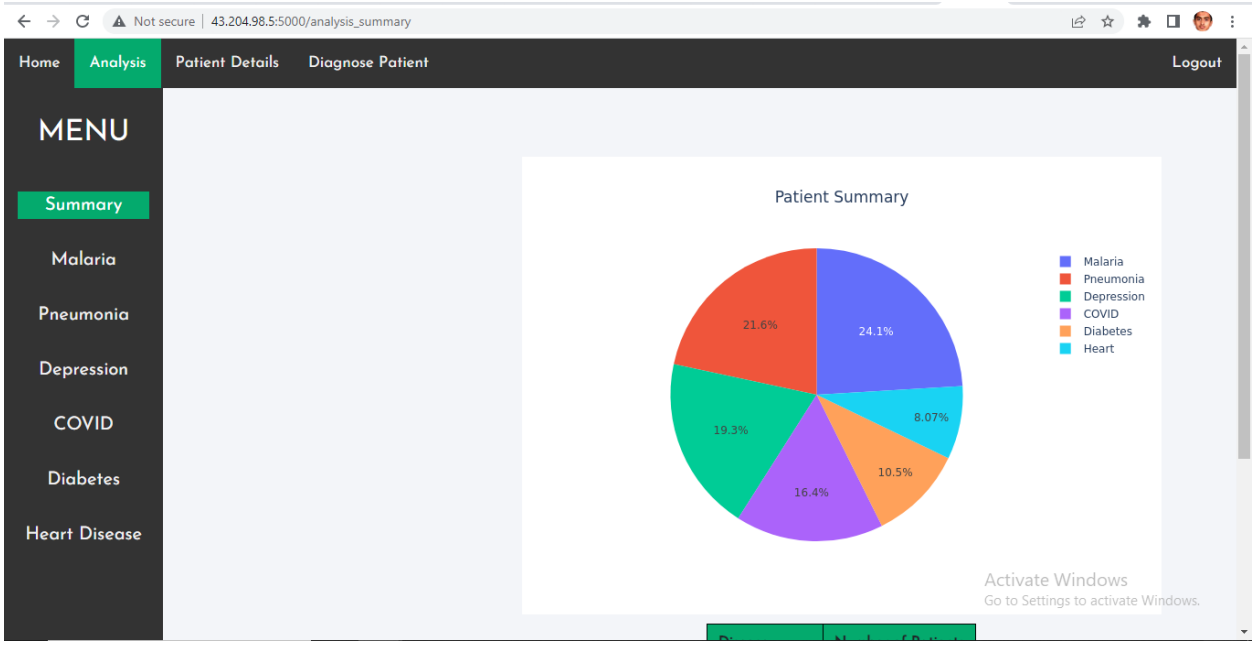
Login page



Home page



Analysis page

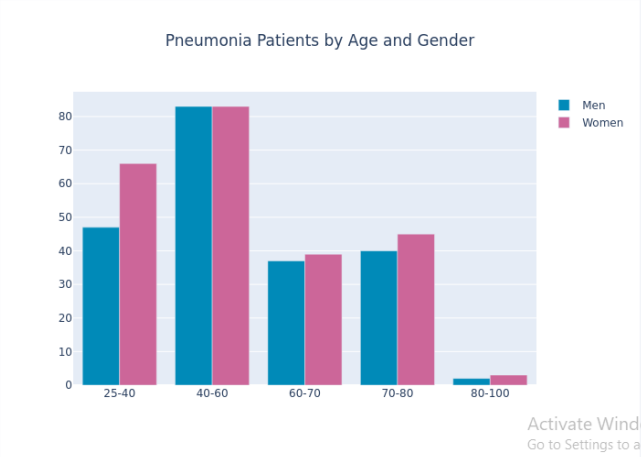


Not secure | 43.204.98.5:5000/analysis_pneumonia

Home | Analysis | Patient Details | Diagnose Patient | Logout

MENU

- Summary
- Malaria
- Pneumonia
- Depression
- COVID
- Diabetes
- Heart Disease

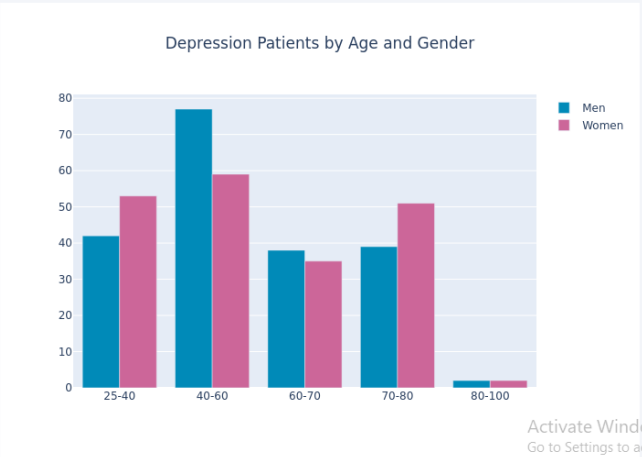


Not secure | 43.204.98.5:5000/analysis_depression

Home | Analysis | Patient Details | Diagnose Patient | Logout

MENU

- Summary
- Malaria
- Pneumonia
- Depression
- COVID
- Diabetes
- Heart Disease

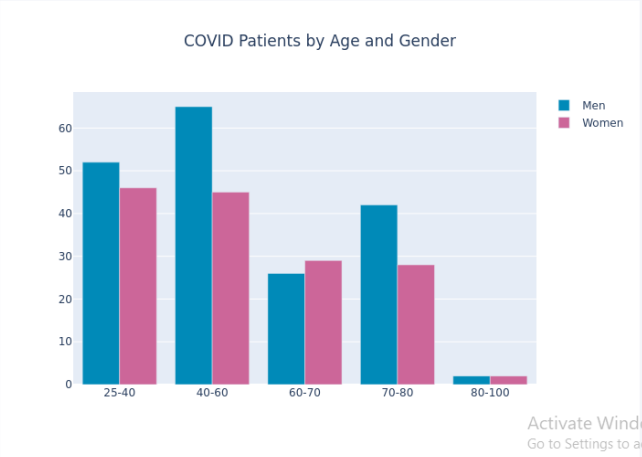


Not secure | 43.204.98.5:5000/analysis_covid

Home | Analysis | Patient Details | Diagnose Patient | Logout

MENU

- Summary
- Malaria
- Pneumonia
- Depression
- COVID
- Diabetes
- Heart Disease

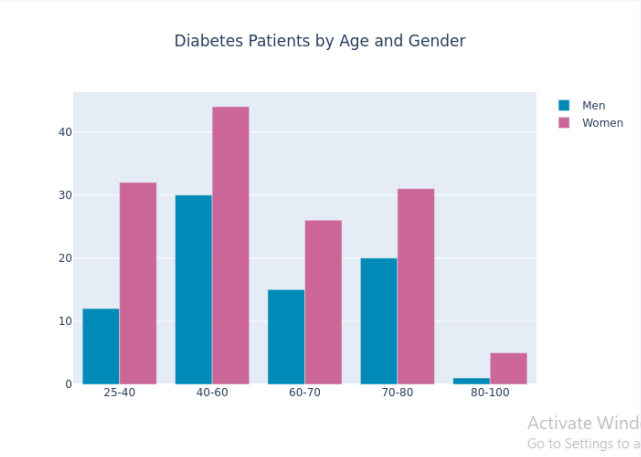


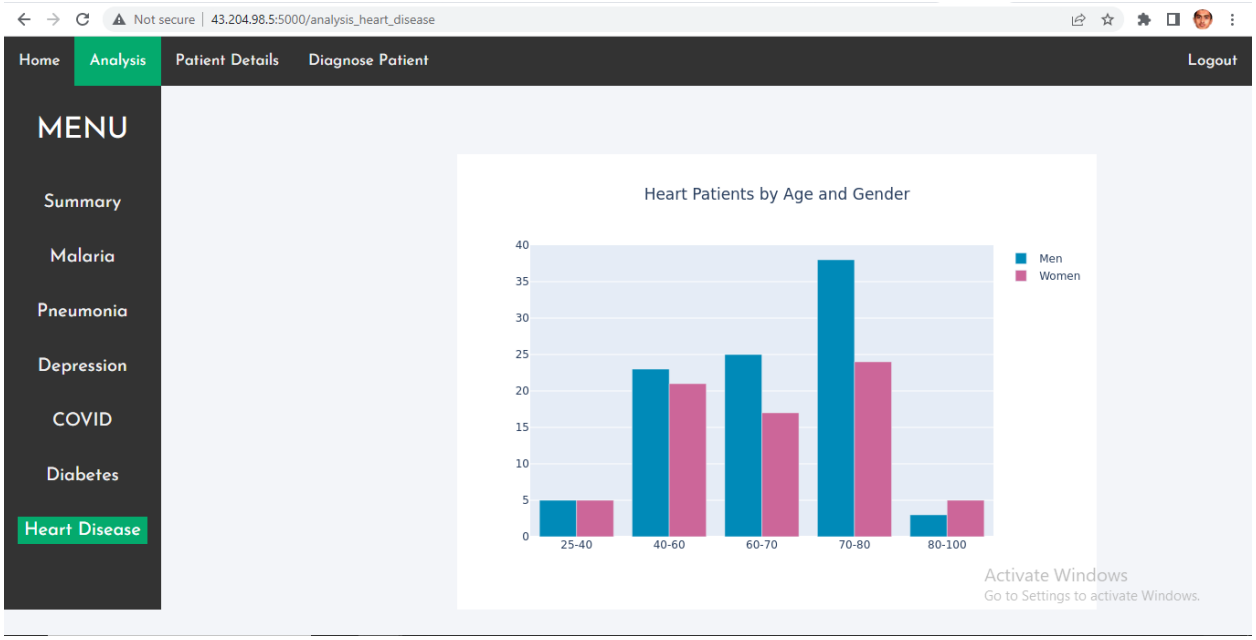
Not secure | 43.204.98.5:5000/analysis_diabetes

Home | Analysis | Patient Details | Diagnose Patient | Logout

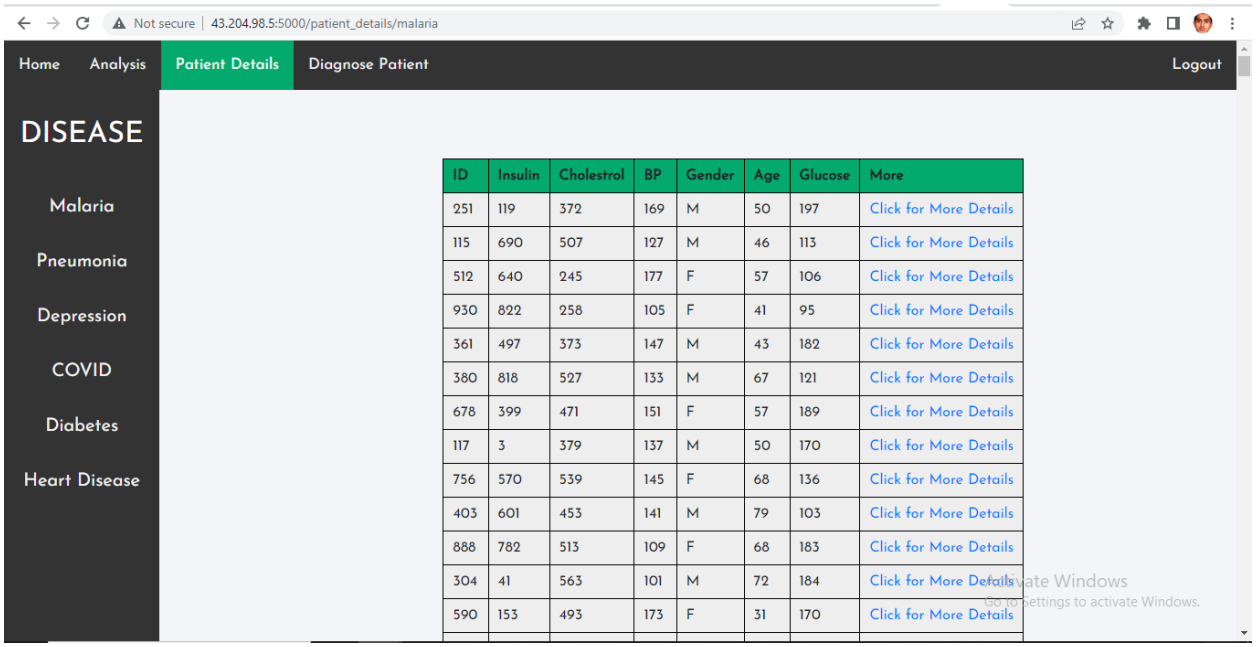
MENU

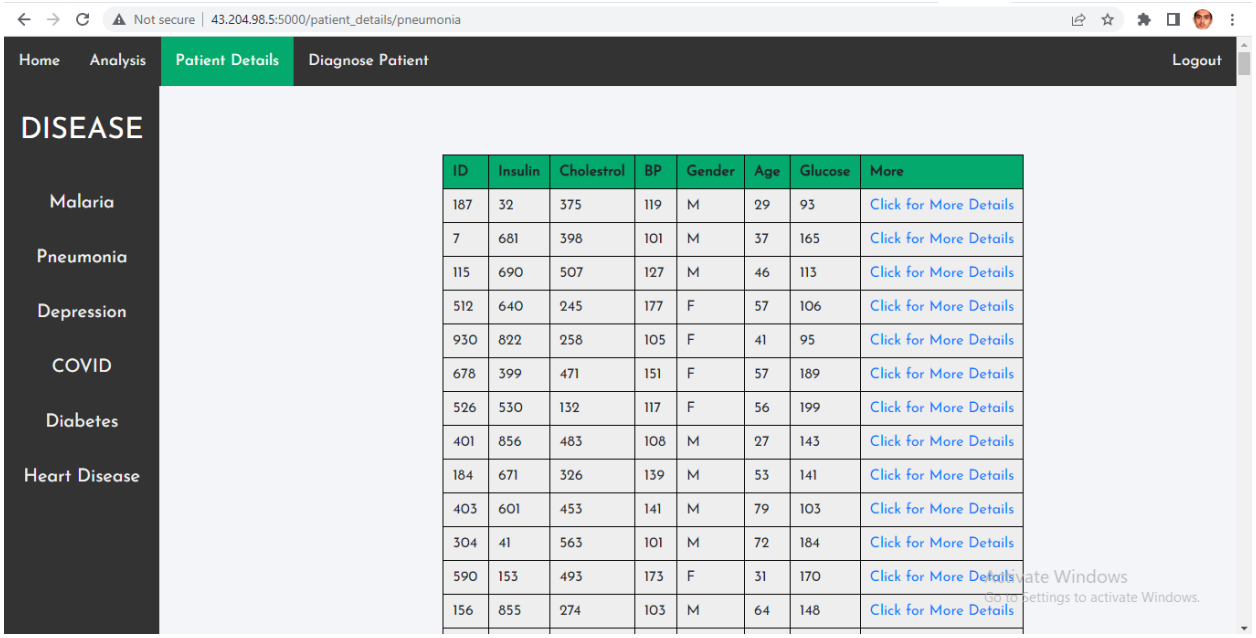
- Summary
- Malaria
- Pneumonia
- Depression
- COVID
- Diabetes
- Heart Disease

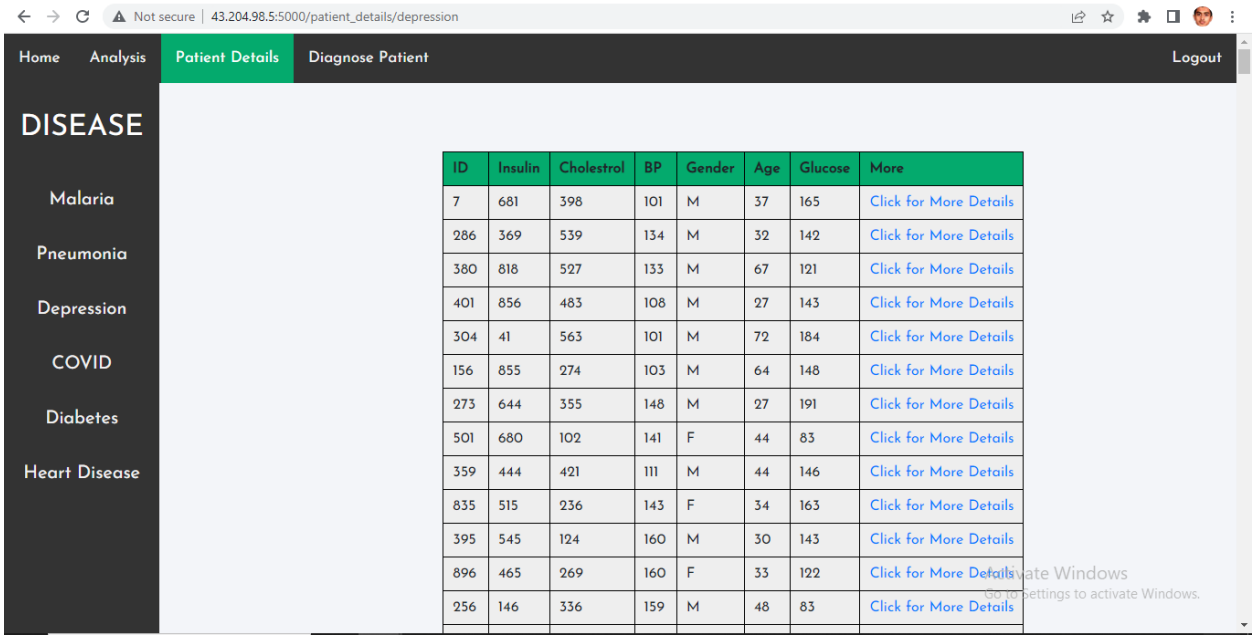


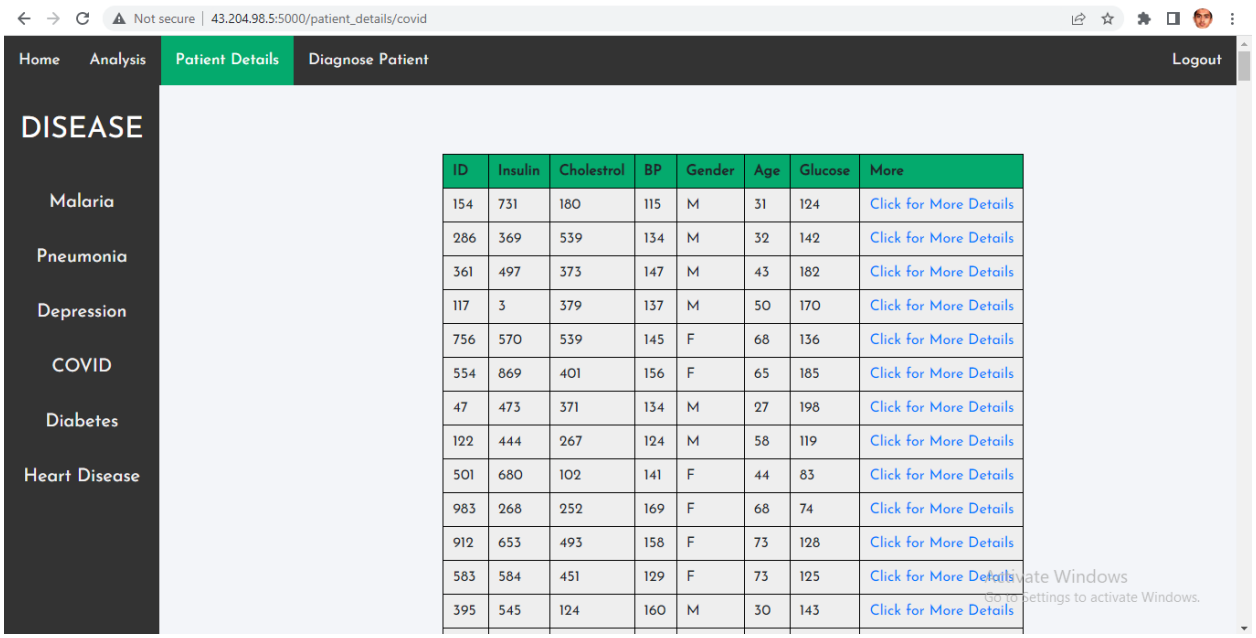


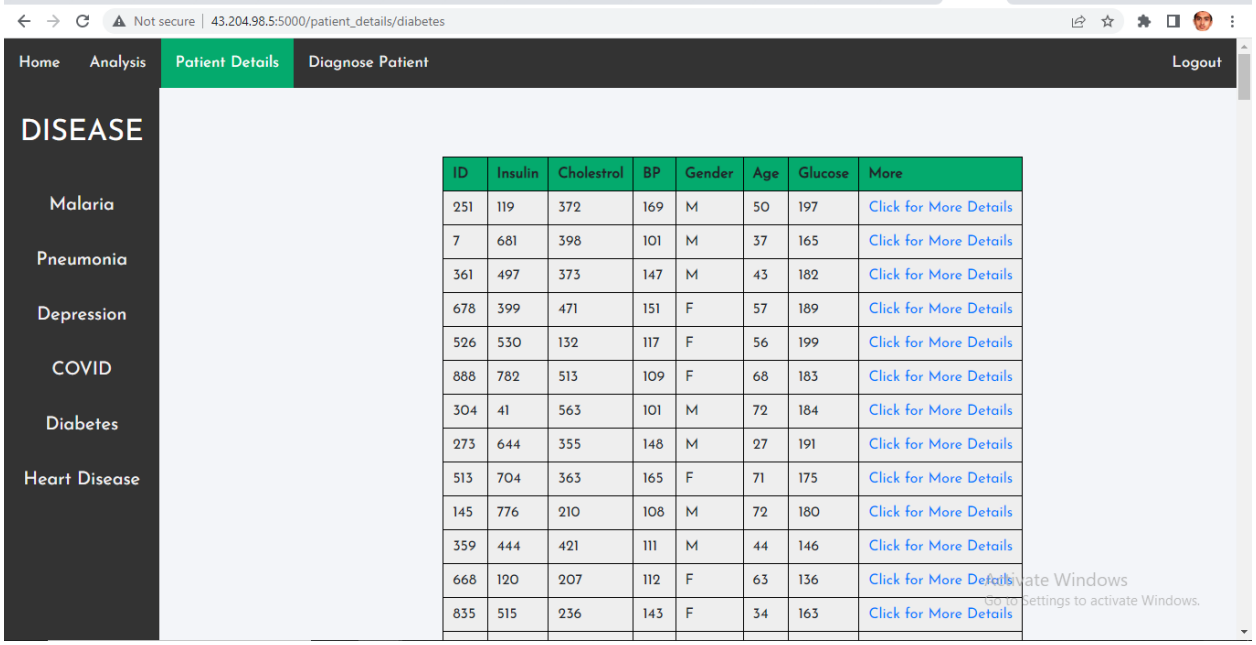
Patient details

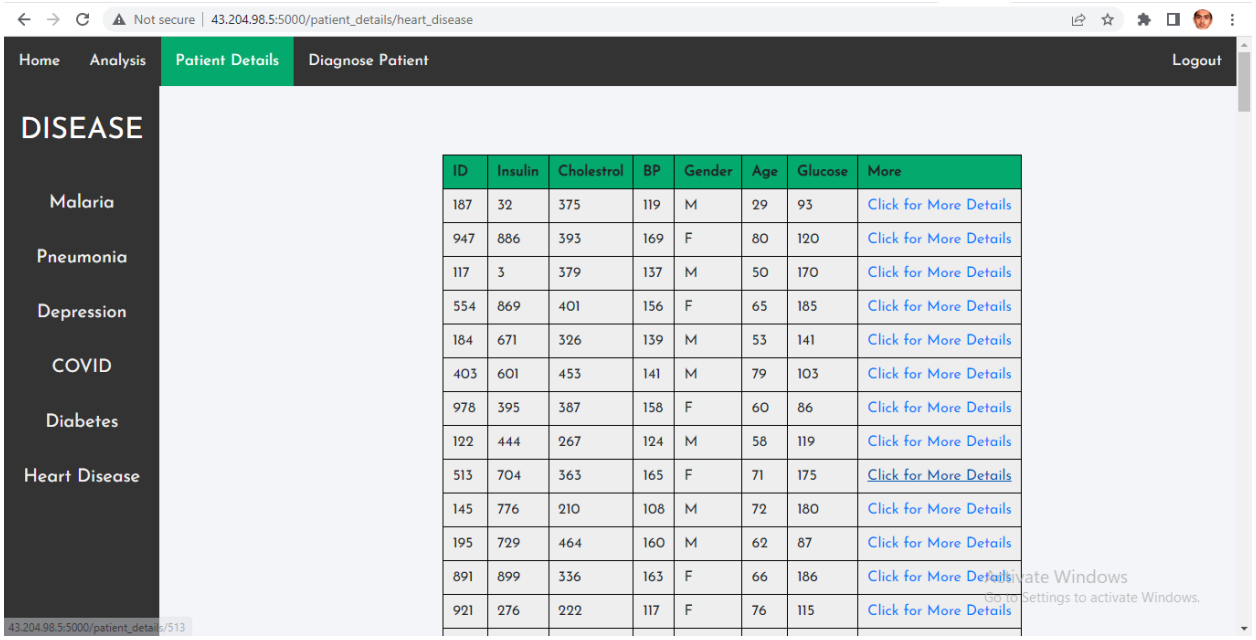


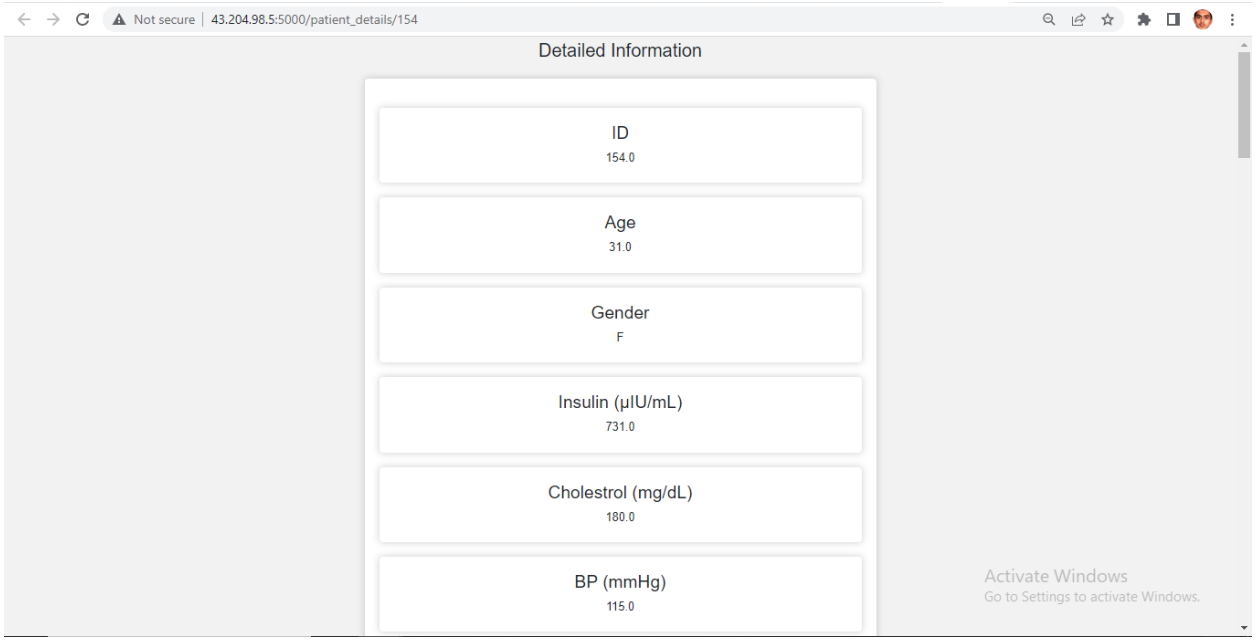

Not secure | 43.204.98.5:5000/patient_details/pneumonia
Home
Analysis
Patient Details
Diagnose Patient
Logout
DISEASE
Malaria
Pneumonia
Depression
COVID
Diabetes
Heart Disease
ID | Insulin | Cholestrol | BP | Gender | Age | Glucose | More
187 | 32 | 375 | 119 | M | 29 | 93 | Click for More Details
7 | 681 | 398 | 101 | M | 37 | 165 | Click for More Details
115 | 690 | 507 | 127 | M | 46 | 113 | Click for More Details
512 | 640 | 245 | 177 | F | 57 | 106 | Click for More Details
930 | 822 | 258 | 105 | F | 41 | 95 | Click for More Details
678 | 399 | 471 | 151 | F | 57 | 189 | Click for More Details
526 | 530 | 132 | 117 | F | 56 | 199 | Click for More Details
401 | 856 | 483 | 108 | M | 27 | 143 | Click for More Details
184 | 671 | 326 | 139 | M | 53 | 141 | Click for More Details
403 | 601 | 453 | 141 | M | 79 | 103 | Click for More Details
304 | 41 | 563 | 101 | M | 72 | 184 | Click for More Details
590 | 153 | 493 | 173 | F | 31 | 170 | Click for More Details
156 | 855 | 274 | 103 | M | 64 | 148 | Click for More Details
Activate Windows
Go to Settings to activate Windows.

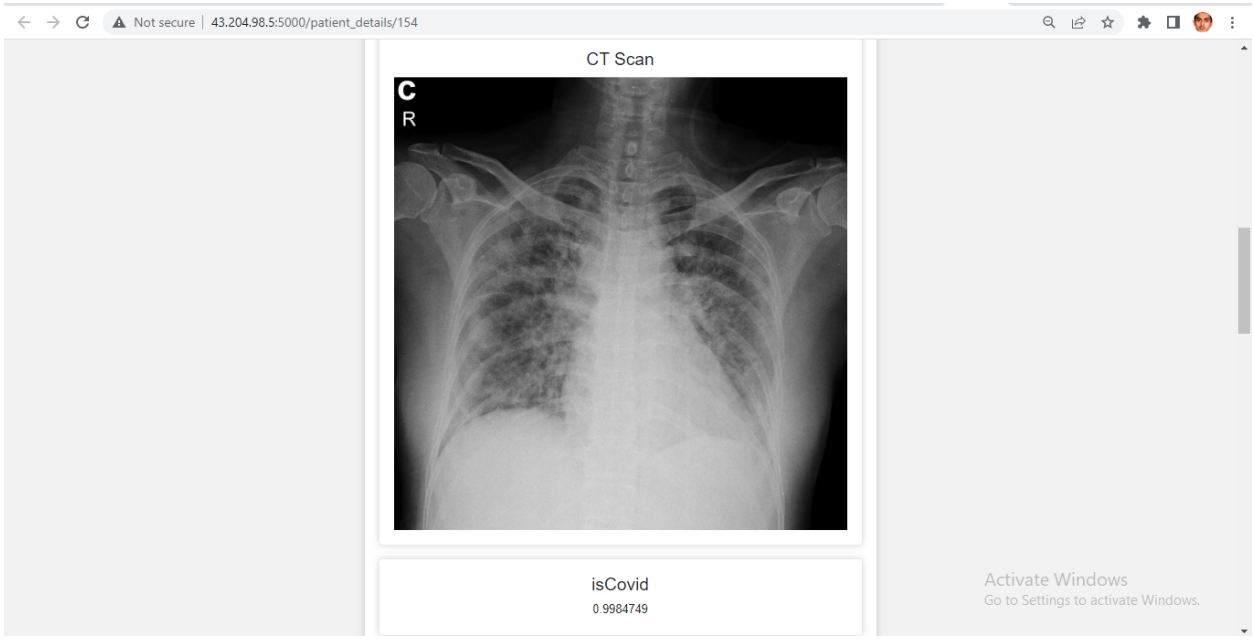

Not secure | 43.204.98.5:5000/patient_details/depression
Home
Analysis
Patient Details
Diagnose Patient
Logout
DISEASE
Malaria
Pneumonia
Depression
COVID
Diabetes
Heart Disease
ID | Insulin | Cholestrol | BP | Gender | Age | Glucose | More
7 | 681 | 398 | 101 | M | 37 | 165 | Click for More Details
286 | 369 | 539 | 134 | M | 32 | 142 | Click for More Details
380 | 818 | 527 | 133 | M | 67 | 121 | Click for More Details
401 | 856 | 483 | 108 | M | 27 | 143 | Click for More Details
304 | 41 | 563 | 101 | M | 72 | 184 | Click for More Details
156 | 855 | 274 | 103 | M | 64 | 148 | Click for More Details
273 | 644 | 355 | 148 | M | 27 | 191 | Click for More Details
501 | 680 | 102 | 141 | F | 44 | 83 | Click for More Details
359 | 444 | 421 | 111 | M | 44 | 146 | Click for More Details
835 | 515 | 236 | 143 | F | 34 | 163 | Click for More Details
395 | 545 | 124 | 160 | M | 30 | 143 | Click for More Details
896 | 465 | 269 | 160 | F | 33 | 122 | Click for More Details
256 | 146 | 336 | 159 | M | 48 | 83 | Click for More Details
Activate Windows
Go to Settings to activate Windows.


Not secure | 43.204.98.5:5000/patient_details/covid
Home
Analysis
Patient Details
Diagnose Patient
Logout
DISEASE
Malaria
Pneumonia
Depression
COVID
Diabetes
Heart Disease
ID Insulin Cholestrol BP Gender Age Glucose More
154 731 180 115 M 31 124 Click for More Details
286 369 539 134 M 32 142 Click for More Details
361 497 373 147 M 43 182 Click for More Details
117 3 379 137 M 50 170 Click for More Details
756 570 539 145 F 68 136 Click for More Details
554 869 401 156 F 65 185 Click for More Details
47 473 371 134 M 27 198 Click for More Details
122 444 267 124 M 58 119 Click for More Details
501 680 102 141 F 44 83 Click for More Details
983 268 252 169 F 68 74 Click for More Details
912 653 493 158 F 73 128 Click for More Details
583 584 451 129 F 73 125 Click for More Details
395 545 124 160 M 30 143 Click for More Details
Activate Windows
Go to Settings to activate Windows.

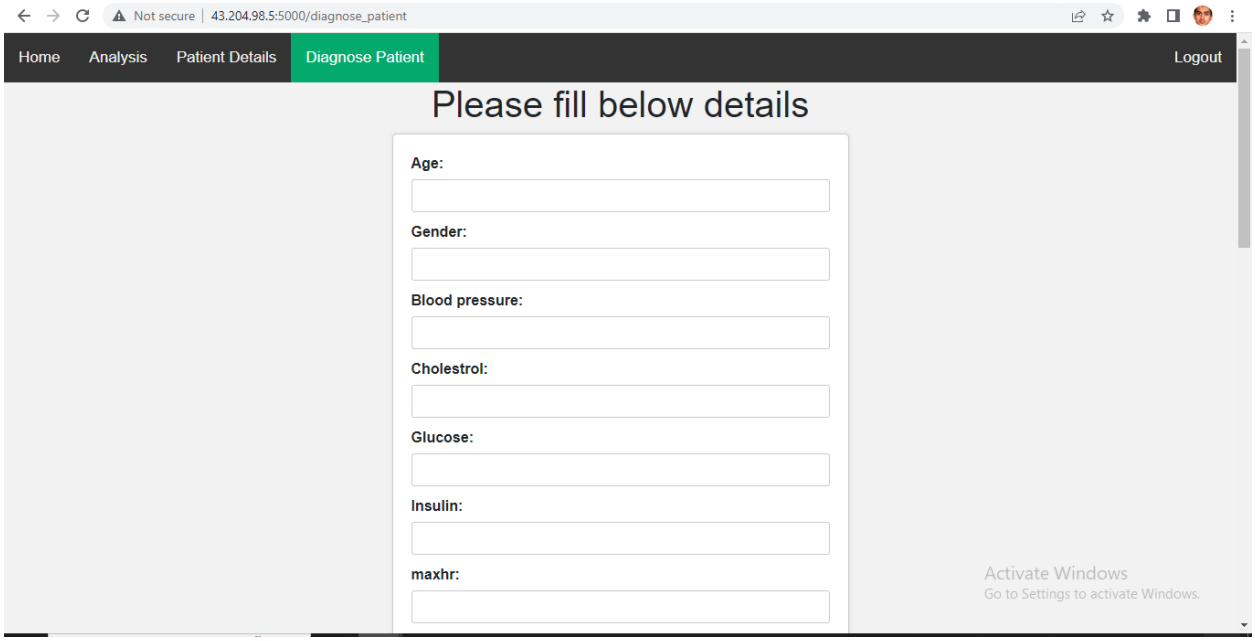

Not secure | 43.204.98.5:5000/patient_details/diabetes
Home
Analysis
Patient Details
Diagnose Patient
Logout
DISEASE
Malaria
Pneumonia
Depression
COVID
Diabetes
Heart Disease
ID Insulin Cholestrol BP Gender Age Glucose More
251 119 372 169 M 50 197 Click for More Details
7 681 398 101 M 37 165 Click for More Details
361 497 373 147 M 43 182 Click for More Details
678 399 471 151 F 57 189 Click for More Details
526 530 132 117 F 56 199 Click for More Details
888 782 513 109 F 68 183 Click for More Details
304 41 563 101 M 72 184 Click for More Details
273 644 355 148 M 27 191 Click for More Details
513 704 363 165 F 71 175 Click for More Details
145 776 210 108 M 72 180 Click for More Details
359 444 421 111 M 44 146 Click for More Details
668 120 207 112 F 63 136 Click for More Details
835 515 236 143 F 34 163 Click for More Details
Activate Windows
Go to Settings to activate Windows.

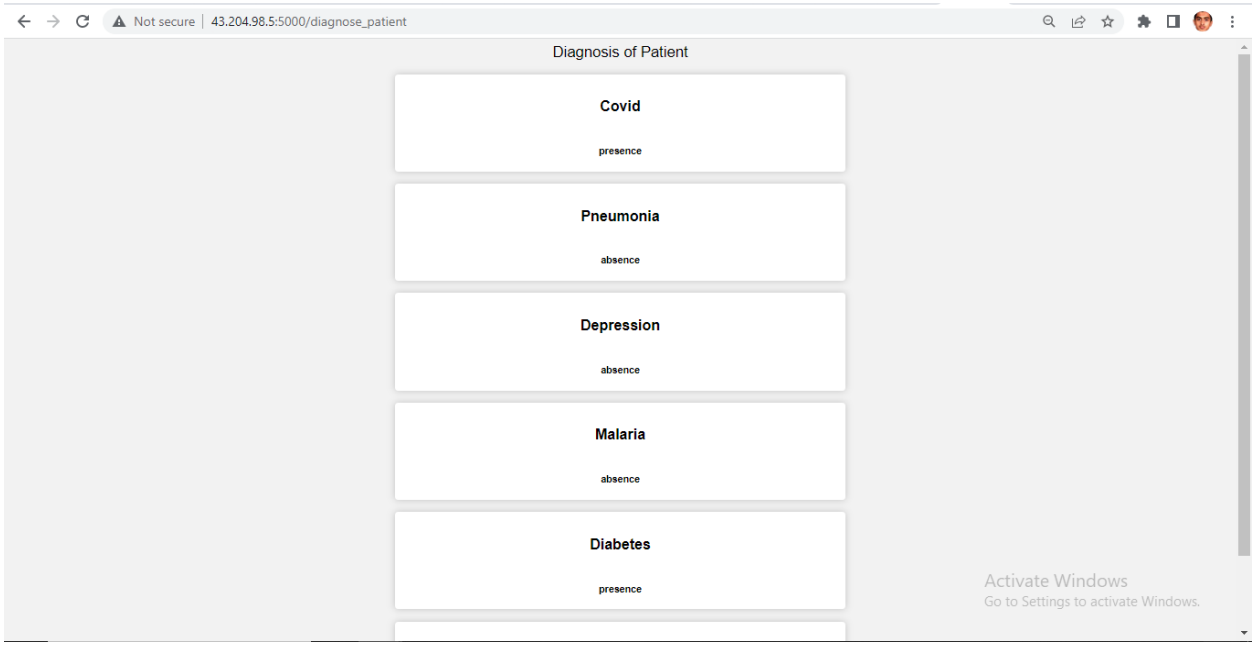

Not secure | 43.204.98.5:5000/patient_details/heart_disease
Home
Analysis
Patient Details
Diagnose Patient
Logout
DISEASE
Malaria
Pneumonia
Depression
COVID
Diabetes
Heart Disease
ID | Insulin | Cholestrol | BP | Gender | Age | Glucose | More
187 | 32 | 375 | 119 | M | 29 | 93 | Click for More Details
947 | 886 | 393 | 169 | F | 80 | 120 | Click for More Details
117 | 3 | 379 | 137 | M | 50 | 170 | Click for More Details
554 | 869 | 401 | 156 | F | 65 | 185 | Click for More Details
184 | 671 | 326 | 139 | M | 53 | 141 | Click for More Details
403 | 601 | 453 | 141 | M | 79 | 103 | Click for More Details
978 | 395 | 387 | 158 | F | 60 | 86 | Click for More Details
122 | 444 | 267 | 124 | M | 58 | 119 | Click for More Details
513 | 704 | 363 | 165 | F | 71 | 175 | Click for More Details
145 | 776 | 210 | 108 | M | 72 | 180 | Click for More Details
195 | 729 | 464 | 160 | M | 62 | 87 | Click for More Details
891 | 899 | 336 | 163 | F | 66 | 186 | Click for More Details
921 | 276 | 222 | 117 | F | 76 | 115 | Click for More Details
Activate Windows
Go to Settings to activate Windows.


Not secure | 43.204.98.5:5000/patient_details/154
Detailed Information
ID
154.0
Age
31.0
Gender
F
Insulin (µIU/mL)
731.0
Cholestrol (mg/dL)
180.0
BP (mmHg)
115.0
Activate Windows
Go to Settings to activate Windows.



Diagnose patient




Not secure | 43.204.98.5:5000/diagnose_patient
Diagnosis of Patient
Covid
presence
Pneumonia
absence
Depression
absence
Malaria
absence
Diabetes
presence
Activate Windows
Go to Settings to activate Windows.

Conclusion

We have successfully developed and deployed a health recommendation system application. We have deployed our application on AWS Cloud which provides highly reliable , scalable and cost effective resources on demand.We have used **AWS S3** for storing images and CSV files related to patients and **AWS Dynamo DB** for storing other patient related information and URL of files stored in S3. We have trained our ML models using **scikit learn** and **tensor flow** libraries which will give predictions of diabetes, heart disease, covid, pneumonia, malaria and mental depression based on the patient information. We have used **Python Flask** library to develop our web application for displaying health recommendation information.Our application performance was good and we didn't experience any issues while accessing our application. In our current application model training, testing , web server are all deployed on a single EC2 instance. However for practical purposes it is better to launch a separate EC2 instance for ML model training , testing and web server. If the dataset size is huge, it would also be better to use Spark ML library on Spark cluster for ML model training and testing. Irrespective of the processing/computation considerations, our storage solution implemented using AWS S3 and AWS DynamoDB is highly scalable and efficient and can work with any processing environment like Mapreduce or Spark.