

General Disease Prediction Using ML

A

Report submitted

in the partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology

in

Computer Science and Engineering

by

Adarsh Gupta: 1900520109001

Jyoti Maurya: 1900520109004

Zainab Israr: 1900520109006

Under the guidance of

Mr. Natthan Singh

Dr. Aditi Sharma



Department of Computer Science and Engineering
Institute of Engineering and Technology, Lucknow
Dr. A.P.J. Abdul Kalam Technical University,
Lucknow, Uttar Pradesh
2021-22

CONTENTS

DECLARATION.....	3
CERTIFICATE.....	4
ACKNOWLEDGMENT	5
ABSTRACT.....	6
LIST OF FIGURES.....	7
LIST OF TABLES.....	8
1. INTRODUCTION.....	9
2. LITERATURE REVIEW.....	10
3. METHODOLOGY.....	11
3.1 PROPOSED MODAL.....	12
3.1.1 DECISION TREE ALGORITHM.....	12
3.1.2 NAIVE BAYE’S CLASSIFIER.....	13
3.1.3 RANDOM FOREST ALGORITHM.....	14
3.2 REQUIREMENT ANALYSIS.....	15
4 SYSTEM DESIGN.....	16
4.1 SYSTEM ARCHITECTURE.....	16
4.2 DATA FLOW DIAGRAM.....	16
4.3 SEQUENCE DIAGRAM.....	17
4.4 USE CASE DIAGRAM.....	18
4.5 ACTIVITY DIAGRAM.....	19
5 IMPLEMENTATION.....	20
6 TESTING.....	21-22
7 TOOLS AND TECHNOLOGY USED.....	23-24
8 RESULTS.....	25-29
9 CONCLUSION AND FUTURE ENHANCEMENT.....	30
9.1 CONCLUSION.....	30
9.2 FUTURE ENHANCEMENT.....	30
10 REFERENCE.....	31

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our belief and knowledge, it contains no material previously published or written by another person or material which to a substantial error has been accepted for the award of any degree or diploma of university or other institute of higher learning, except where the acknowledgement has been made in the text. The project has not been submitted by us at any other institute for requirement of any other degree.

Date: 29/05/2022

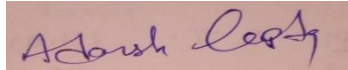
Submitted by: -

(1) Name: **Adarsh Gupta**

Roll No.: 1900520109001

Branch: CSE

Signature:



(2) Name: **Jyoti**

Roll No.: 1900520109004

Branch: CSE

Signature:

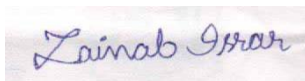


(3) Name: **Zainab Israr Ansari**

Roll No.: 1900520109006

Branch: CSE

Signature:



CERTIFICATE

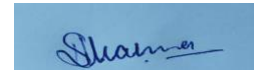
This is to certify that the project report entitled “General Disease Prediction Using ML” presented by Adarsh Gupta, Jyoti and Zainab Israr Ansari in the partial fulfillment for the award of Bachelor of Technology in Computer Science and Engineering, is a record of work carried out by them under my supervision and guidance at the Department of Computer Science and Engineering at Institute of Engineering and Technology, Lucknow.

It is also certified that this project has not been submitted at any other Institute for the award of any other degrees to the best of my knowledge.



Mr. Natthan Singh

Department of Computer Science and Engineering
Institute of Engineering and Technology, Lucknow



Dr. Aditi Sharma

Department of Computer Science and Engineering
Institute of Engineering and Technology, Lucknow

ACKNOWLEDGEMENT

We bow in reverence to God who gave us the required zeal for making the project. We are deeply indebted to our Head of the Department **Dr. Divakar Singh Yadav**. We owe, department gratitude to **Mr. Natthan Singh** and **Dr. Aditi Sharma** of IET Lucknow for their valuable guidance and suggestions and treatment for giving helpful guidelines for the project. We also thanks to Project Committee for their Proper guidance, support, and suggestion to us time to time.

We feel highly proud in expressing our coordinate thanks to respected parents, brothers, and friends who have helped us in this project.

Adarsh Gupta: 1900520109001

Jyoti Maurya: 1900520109004

Zainab Israr Ansari: 1900520109006

ABSTRACT

General Disease Prediction using Machine Learning is a system that predicts the disease on the basis of symptoms that the user enters into the system and then it provide the accurate result based on those symptoms that are entered. This system works as an initial phase diagnosis of disease when the user is not much serious or not able to go hospital or not sure about disease and he/she wants to know disease. In today's world health industry plays critical role in curing the diseases that the patient is suffering from. So this system will be helpful for the health industry. It is useful for those person who doesn't want to go to the hospital or any other clinics, then only by entering the symptoms, he/she can get to know the disease that he/she has been suffering from and also the healthcare sectors can also use this system and get benefit as only by asking the symptoms of the patient and entering it in the system and within seconds the system will tell the exact and accurate diseases. This General Disease Prediction Using Machine Learning project is a web application, which is based on Django framework and python language is used for implementation of Machine Learning Model and taking reference from previously available dataset of hospitals. Using this dataset system will predict the accurate disease of each and every patient.

LIST OF FIGURES

Fig 3.1 Decision Tree Example.....	12
Fig 3.2 Naïve Baye's Example.....	13
Fig 3.3 Random Forest Example.....	14
Fig 4.1 System Architecture.....	16
Fig 4.2 Data Flow Diagram (Level-0).....	17
Fig 4.3 Data Flow Diagram (Level-1).....	17
Fig 4.4 Sequence Diagram.....	18
Fig 4.5 Use Case Diagram.....	18
Fig 4.6 Activity Diagram.....	19
Fig 8.1 Training Data.....	25
Fig 8.2 Testing Data.....	25
Fig 8.3 Registration Page.....	26
Fig 8.4 Login Page.....	26
Fig 8.5 Login Page showing invalid Email id/ Password.....	27
Fig 8.5 Dashboard.....	27
Fig 8.6 Dashboard with chosen Symptoms.....	28
Fig 8.7 Dashboard with predicted output.....	28
Fig 8.8 Past record of patient.....	29

LIST OF TABLES

Table 6.1 Test Case for User Registration.....	21
Table 6.2 Test Case for User Login.....	22
Table 6.3 Test Case for Prediction of Result.....	22

1.INTRODUCTION

Nowadays, when anyone suffers from any health-related issues, then the person has to visit a doctor which takes time and it is costly too. It has been observed mostly that every 2 months, over 70% of the population in India tends to general body disease like viral fever, cold and cough, etc. Since a lot of people don't realize that the symptoms of these regular ailments may be symptoms of something more dangerous, 25% of the population succumbs to death due to ignorance of the early-stage symptoms. Thus, the identification of the disease in the initial stages is important for the prevention of any unwarranted casualties. The medical system is mainly devoted to specific areas, known diseases and is insufficient to identify and accurately diseases based on early-stage symptoms. Also if the user/patient is not able to reach hospital and is unable to consult a doctor it may be difficult for the user/patient as the disease can not be identified. So, if the above problems are resolved by using an automated application that can be less time-consuming as well as money, it will be very helpful and easier for the user.

The “General Disease Prediction” will be a web application that will give predictions on the basis of symptoms provided by the user. It will help the user/patient to identify the disease on the basis of symptoms that he/she will enter into the system and provide accurate results on the basis of those symptoms.

The application is designed in such a way that it eliminates the errors as much as possible, while entering the information. It will give suggestions while entering the information. This application would be very easy to use as there is no formal knowledge required for a person to use this application. It is a user-friendly system, which leads to an error-free, reliable, safe, secure and quick prediction system. It helps the users to know their disease and they can take some precautions.

2. LITERATURE REVIEW

- Pahulpreet Singh Kohli, Shriya Arora [2018], proposed their work as “Application of Machine Learning in Disease Prediction”. In this paper, the accuracy reaches 87.1% using logistic Regression and 85.71% using Support Vector Machine [1].
- Dhiraj Dahiwade, Prof. Gajanan Patle, Prof Ektaa Meshram [2019], presented their work as “DESIGNING DISEASE PREDICTION MODEL USING MACHINE LEARNING Approach”. There they used two algorithms (K-Nearest and CNN) to work on their project. It was found out that the accuracy of CNN was 84.5% which was more than KNN algorithm. The memory and time requirement was found to be more in KNN as compared to CNN [2].
- Nikita Pawar, Pooja Sonawane, Narendra Gawai[2020] made a web-based application named “DECISION TREE BASED HEALTH PREDICTION SYSTEM”. In this application, the user provides the symptoms he/she is facing and checks for further symptoms. Accordingly the system checks the InfoBase and extracts the data from it and predicts the accurate disease that the person is affected by. In this method it provide multiple symptoms option through which the patient to search for the symptom he/she is facing. multiple symptom options so that the patient could search for every possible If the symptoms that the user had entered doesn't match any disease in the database, then it displays the disease that the user could probably have related to his/her symptoms. If the system isn't ready to provide suitable results, it urges users to travel for biopsy, x-ray, or CT scan [3].
- Nishant Yede , Ritik Koul , Chetan Harde , Kumar Gaurav , Prof. C.S.Pagar [2021] proposed their work as “GENERAL DISEASE PREDICTION BASED ON SYMPTOMS PROVIDED BY PATIENT”. They have applied three algorithms namely Naive Baye’s Classifiers, Random forest and Decision tree. It was found out that this model have better accuracy as compared to the existing ones. When the result of different algorithms are compared, the accuracy of the proposed algorithm was found to be 94.8% having regular speed which is faster as compared to the unimodal disease risk prediction algorithm [4].

3. METHODOLOGY

Since we all know that mankind is involved so much in this competitive era of economic development that he/she is not much concerned about their health. So, this project General Disease Prediction using Machine Learning is designed to identify the type of disease in earlier phase. It is implemented using python and django framework and is converted into a web application where the users have to register first to get access to the system.

The system has user module only. User is a person who wants to check for the disease based on the symptoms. If the user is already registered then he/she can log in to the system otherwise he/she has to firstly register in the system. The user can register into the system with username, email-id, date of birth and password. All these details of the user are stored in the database. After successful registration, the user can login into the system. The login can be done using email address and password which the user has provided during the time of registration. Authentication is done by the system for each user. If the details entered regarding email address/password are incorrect then the error message will prompt stating that incorrect email address/password. So, correct email address and password both is essential to login into the system.

After logging in the user will reach to the dashboard page where he/she can perform following functionalities:

- a. Entering Symptoms
- b. Disease Prediction
- c. Previous Disease

Step 1: Entering Symptoms: After logging into the system the main page will appear on which drop-down menu is given to select the symptoms. The user can select the symptoms from the list given in the form of drop down menu. To have better accuracy of the result, the user is required to enter more than three symptoms. In our project, three algorithms of machine learning are used for prediction of result. These algorithms are Decision Tree, Random Forest and Naïve Baye's are there for prediction. The system allows the patient to enter the symptoms and based on those symptoms, the system will predict the disease.

Step 2: Disease prediction: After entering all the symptoms the user needs to press the predict button. The result of the prediction will be outputted on the screen. We have used 3 algorithms to predict the type of disease and in final the result of the algorithms having highest accuracy will be shown to the user.

The main purpose behind implementing this method is that if the user does not have experience in the medical profession and want to know about their health conditions, he/she can quickly find it without help of technical or medical person. We have tried to design the user interface to be as much interactive so that it is user friendly.

Step 3: Previous Disease: The user can view the previous record of the diseases that he has been through.

3.1. Proposed Modal

Machine learning is the important principle in which the system provides more precise predictions. Disease Prediction can be implemented using different techniques like Support Vector Machine, Neural networks, decision tree and Naïve Bay's algorithm, etc. But, In our project we have used three algorithms namely Decision Tree, Naïve Baye's and Random Forest.

3.1.1 Decision tree algorithm

- It is a type of Supervised Machine Learning in which the data is splitted continuously on the basis of certain parameters.
- It has a tree-like structure having two entities: leaves and decision nodes. The leaves represent decisions or the final outcomes of the problem and the decision nodes are where the data is split further nodes.
- Decision trees solve two types of problem that is classification (it is categorical data type) and Regression (it is continuous data type).
- It has a very good accuracy as compared to other algorithms.
- The accuracy of decision tree is not affected by the presence of redundant attributes.

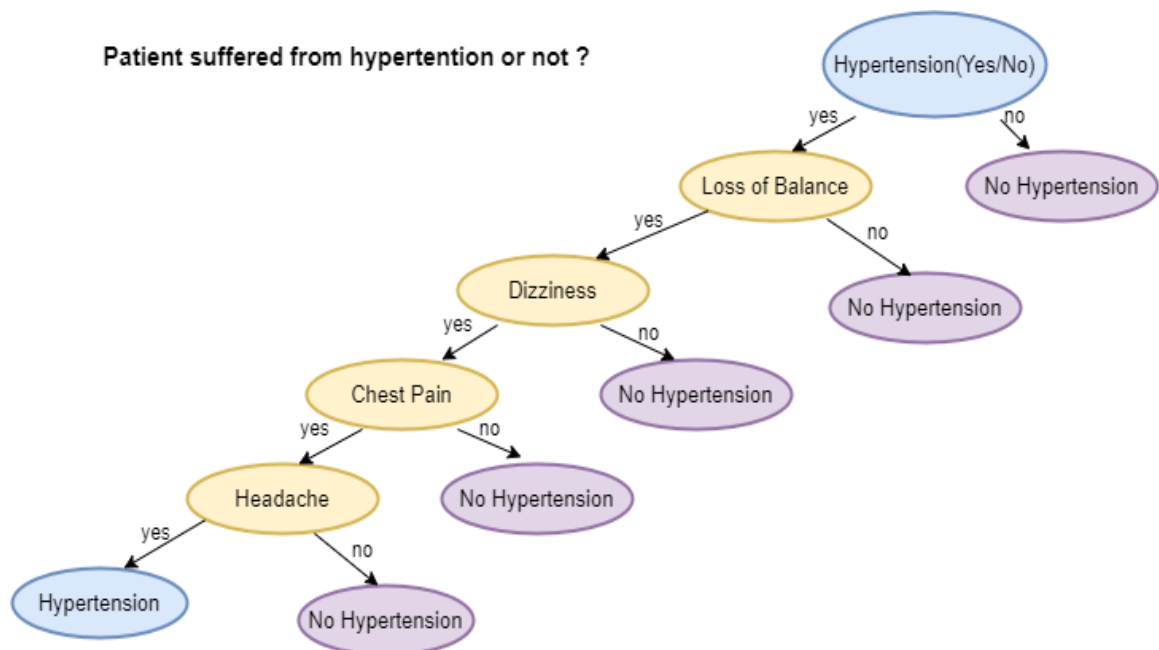


Fig 3.1: Decision Tree Example

3.1.2 Naïve Bay's classifier

- Naïve Baye's is a supervised learning Algorithm, which is used for prediction problem that have categorical level data.
- On the basis of training set and the values in a classifying attribute, it classifies the class data and uses it in classifying new data that is used for training purpose.
- It uses Baye's theorem for classification of problems that uses probabilistic approach.
- It works on both continuous value attributes and categorical type dataset.
- In this algorithm, firstly find the probability of an event occurring when the probability of another event that has already occurred is given. It is based on following equation:

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

- The working of this algorithm with the dataset is described as:
- It divides the given dataset into two parts: response vector and feature matrix.
- Feature matrix contains vectors which represent rows of the dataset and the value of dependent features of dataset will be stored and represented by every vector of matrix.
- Response vector consists of values of class variable for every row of feature matrix.
- The fundamental assumption of naïve bay's classifier is that equal and independent contribution is made by each.

Patient has Hypertension or not ?

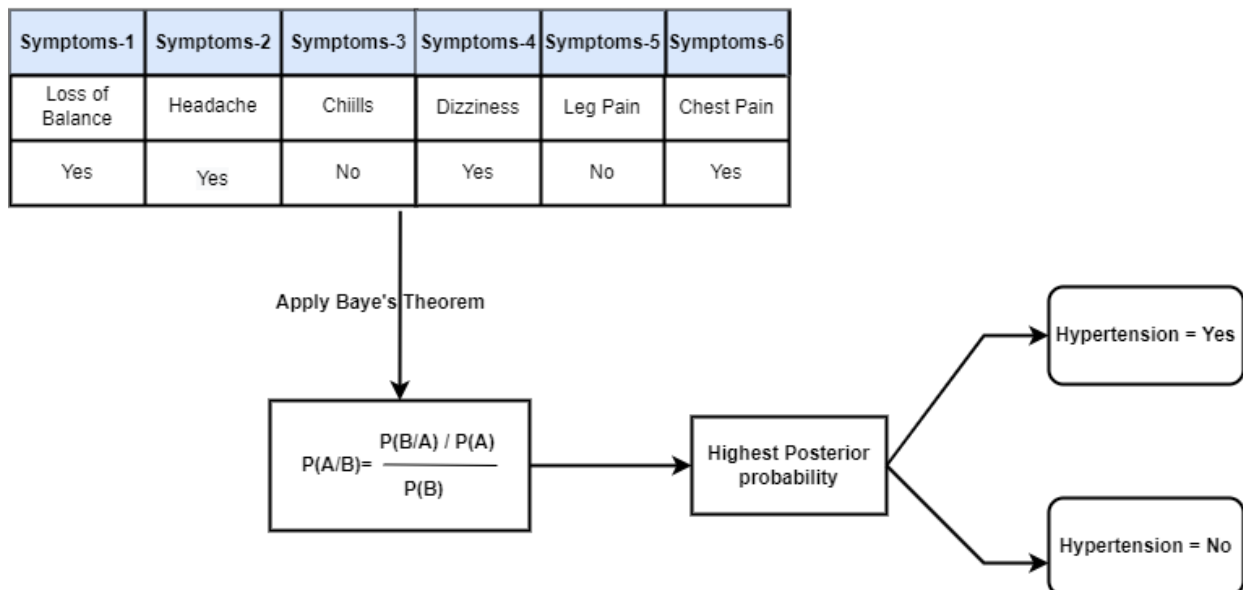


Fig: 3.2: Naïve Baye's Example

3.1.3 Random Forest Algorithm

- It is a supervised machine learning technique that can be used for both types of problems namely: Classification and Regression.
- It is based on the concept of ensemble learning that have several decision tree models that have individual outcome.
- In random forest there are multiple decision trees on different subsets of the training dataset and to improve the predictive accuracy of the dataset and majority voting outcome is taken.
- Random forest classifier does not depend on single decision tree, instead it evaluates the prediction from every decision tree and on the basis of majority voting of predictions, it predicts the final output based on that.
- It uses Gini Index for determining the final class in each tree.
- Large number of decision trees in the Random Forest Algorithm is used to achieve better accuracy and to prevent from the problem of overfitting.

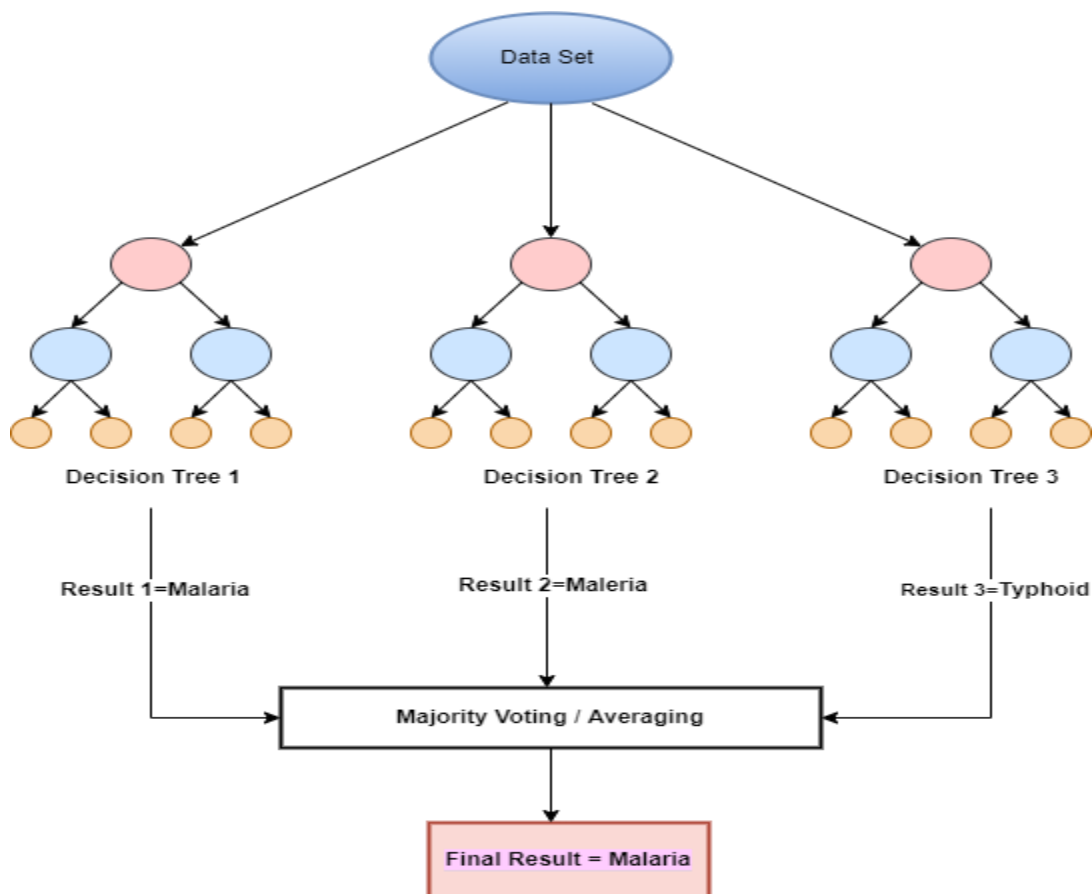


Fig 3.3: Random Forest Algorithm Example

3.2. Requirement Analysis

Tools/Platform/Software and Hardware Specification:

Hardware Requirement:

- ❖ System : Pentium 4, Intel Core i3, i5, i7
- ❖ RAM : 512 Mb or above
- ❖ Hard Disk : 10 Gb or above
- ❖ Input Device : Keyboard and Mouse
- ❖ Output Device : Monitor or PC

Software Requirement:

- ❖ Operating system : Windows XP, 7, 10 or Higher Versions
- ❖ Browser : Google Chrome, Firefox, IE 10.0 or later
- ❖ Database : MySql

Developer

- ❖ Database : MySql
- ❖ IDE : PyCharm
- ❖ Documentation Tool: MS Word, MS Power-point
- ❖ Front End : HTML, CSS, JavaScript
- ❖ Framework : django
- ❖ Back End : Xamp
- ❖ Programming Language: Python

4. SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE

In the architecture of our project general disease prediction using machine learning have different datasets on which comparison is done with the symptoms that the user entered and predicts it. The datasets are converted into the small data sets that is training dataset and testing dataset, after that the data which is classified is processed into the machine learning technologies. The disease prediction model using all the input data from the user and overall processed data compares the relation and predict the final disease accordingly. After training phase, model uses testing dataset for evaluating the prediction model.

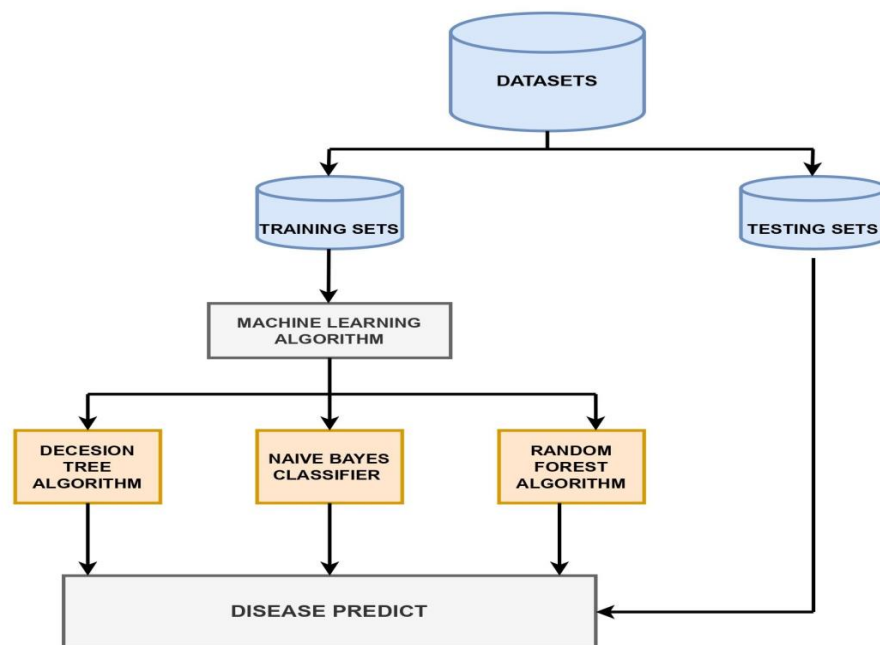


Fig 4.1: System Architecture

4.2 DATA FLOW DIAGRAM

This dataflow diagram of the project “General disease prediction using machine learning” shows how the data flows from one step to another in the model, like the patient enters the details and symptoms into the system. This data goes into the system, and a comparison is made with the prediction model and if true it predicts the appropriate result.

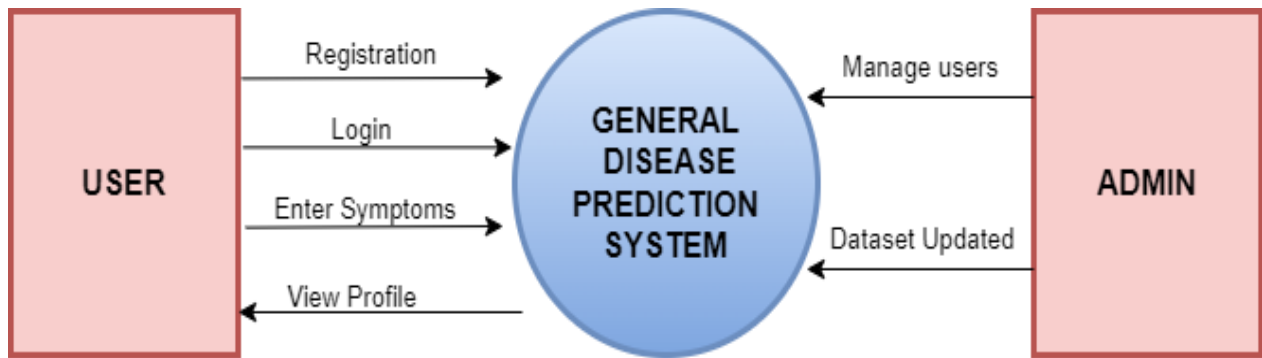


Fig 4.2: DFD(Level-0)

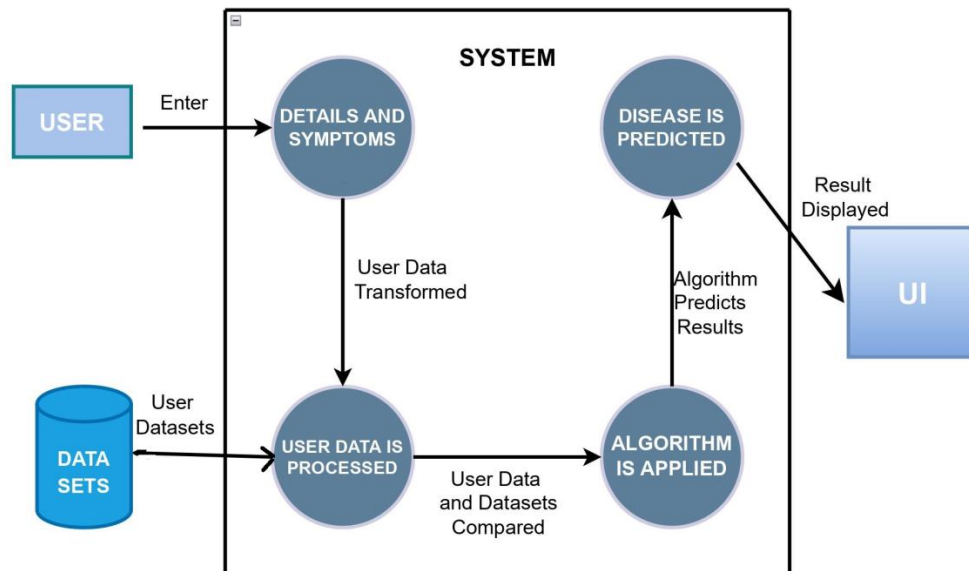


Fig 4.3: Data Flow Diagram (Level-1)

4.3 SEQUENCE DIAGRAM

The sequence diagram represents how the model flows from one step to another step, like the user enters into the system through login then symptoms of that he/she is having, that enters into the system, then it is compared with the predictive model and if the result turns out to be true then it predicts the results otherwise it shows the details that the user have entered incorrectly and it also shows the precautionary steps for the user to follow. Here the sequences of all the entities are linked to each other where the user gets started with the system.

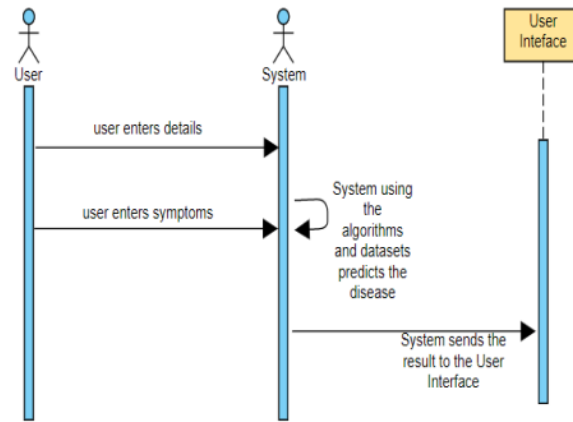
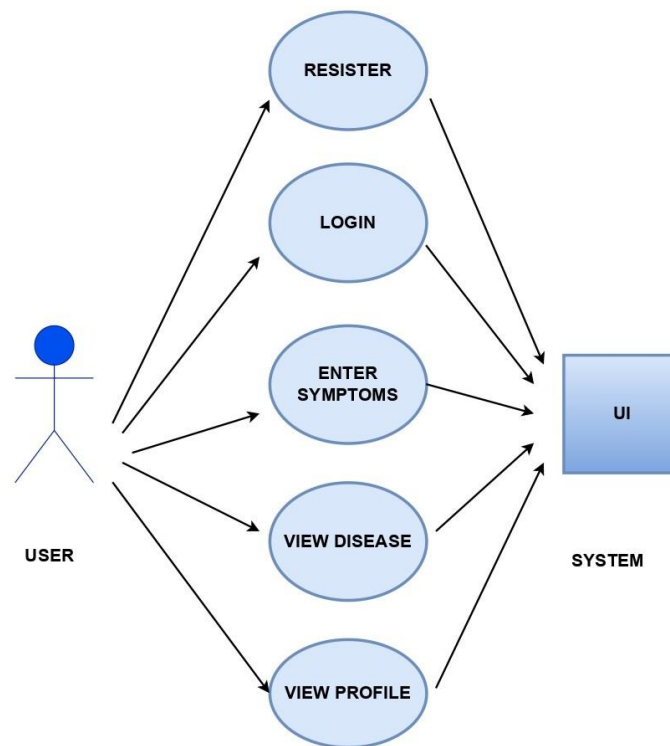


Fig 4.4: Sequence Diagram

4.4 USE CASE DIAGRAM

This use case diagram shows the description of the functionality of a system from user perspective like the user enters the symptoms that goes into the system, and a comparison is made with the prediction model and it predicts the output.



USE CASE DIAGRAM

Fig 4.5: Use Case Diagram

4.4 ACTIVITY DIAGRAM

Activity Diagram is a flowchart that shows the flow from one activity to another activity. The operation of the system is described as an activity. From one activity to other activity, the flow of control is drawn. In this activity diagram, the activity starts from user where the user firstly registers into the system to get access of the system then login using the credentials and then the credentials are verified with the system and if its verified, then the user proceeds to the main dashboard page where he/she can perform prediction. Then after entering the symptoms and processing the data from datasets the analysis will happen after that the correct result of the prediction will be displayed.

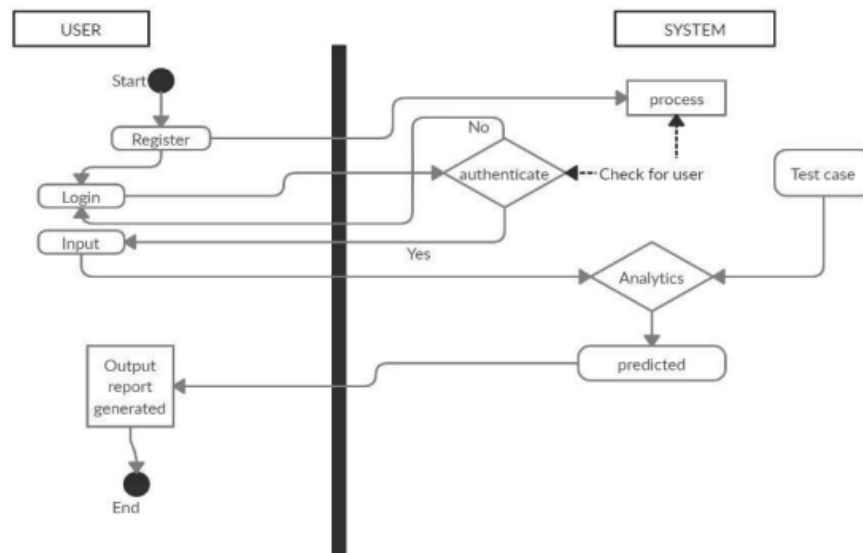


Fig 4.6 Activity Diagram

5. IMPLEMENTATION

The system is designed in such a way that it is user friendly. Each time when the user enters into the system, authentication is required. After which the system provides the result based on the input entered by the user. The implementation of the system is described below:

- Once user opens the system he/she needs to register itself first to get the access to the system, if the user is not registered previously.
- For registration user needs to provide some basic details (Username, Email Address, Date of Birth, and Password) for signup and then the details of user are saved in system.
- Then user is required to login to have a checkup of his/her health.
- For login, users have to provide Email Address and Password.
- While logging in if the user details are wrong then, the system will show a prompt message which states that the user incorrect username or password
- Hence it is important for the user to enter the correct email address as well as password to get access to the system.
- After logging in, the user will be navigated to the dashboard page where he/she can enter the symptoms that the user is having.
- Based on the symptoms the predicted output will be displayed as result.
- We have used several algorithms which will predict the disease and the result of the algorithm having best accuracy will be displayed as output.
- The user needs to enter at least 3 symptoms and maximum 5 symptoms to get the accurate result.
- Data collection and dataset preparation: This involves the collection of information related to medical from different sources like hospitals or other health care industries, after that pre-processing is done on that dataset that removes all the information which is not necessary and negatively affected the output of model and extract important features from dataset using statistical approaches. We have collected the dataset from Kaggle.
- Training the datasets: The General Disease Prediction model is trained on the dataset of diseases to predict accurately. In this system 3 different algorithms were used:
 - Decision Tree Algorithm
 - Naïve Bay's Algorithm
 - Random Forest Algorithm

6. TESTING

6.1 UNIT TESTING

Unit testing is a type of testing in which test cases are designed in such a way that they validate the proper functioning of the internal program logic. The input provided to the program should produce valid outputs. Validation should be done for all the internal code flow and decision. It tests the individual software units of the application. Unit testing is performed before integration and after the individual unit is completed. It is a sort of structural testing as it relies on knowledge of its construction. Testing is performed at the level of component and tests specific processes or application.

6.2 INTEGRATION TESTING

Integration tests is a type of testing in which testing is done on the integrated software components to check whether it runs as a one program or not. It is performed after all the software components are integrated as a single program. Integration testing is done to show that even after the components are individually tested by unit testing, the combination of each component is correct and consistent. The purpose of integration testing is to expose the problems that occur from the combination of components.

6.3 SYSTEM TESTING

System Testing is a type of testing which is performed after the integration of hardware and software system for the verification that the system meets its specified requirements. System Testing confirms that all the code modules work properly as specified, and also the system as a whole performs well on the platform on which it will be deployed.

Description	Input	Output	Remarks
User enters detail for registering in the system	Name, Email Address, Date of Birth, Password,	If the details of the user are correct, the user gets registered. If the details are incorrect, it displays error message.	Test Successful

Table 6.1: Test Case for User Registration

Description	Input	Output	Remarks
User enters detail for Login in the system	Username and Password	If the details entered for login are correct, the user will get logged in into the system and dashboard page will be displayed. If the details entered for login are incorrect then it shows error message.	Test Successful

Table 6.2: Test case for User Login

Description	Input	Output	Remarks
The user will be required to enter minimum 3 and maximum 5 symptoms	Symptoms	If user enters all five symptoms correctly then the accuracy will be high. If less than 3 symptoms are entered, then accuracy will be low.	Test Successful

Table 6.3: Test case for Prediction of Disease

7. TOOLS AND TECHNOLOGY USED

Python

Python is a multi-paradigm programming language. It supports Object-oriented programming and structured programming.

HTML

It is known as Hyper Text Markup Language. It describes the structure of the web page . It provides information in the document by denoting certain text as links, heading, paragraph, list and so on. It is written in the form of tags, surrounded by angle brackets. It also describes some semantics and the appearance of the document. It provides functionality to embed scripting language code which affects the behavior of web browser.

CSS

CSS, an initialize of Cascading Style Sheet, is a simple designing language for the web page. It is used to make the process of making the web page presentable in a simplified manner. The functionalities provided by CSS are one can control the text color, font style, layout and size of columns, display variations for different devices and screen size, space between paragraphs, layout, designs, color and background images.

CSS provides a way through which the presentation of HTML document can be controlled. It is easy to understand and learn. It is used in combination with markup languages like HTML or XHTML.

JAVASCRIPT

It is a text-based programming language which is utilized for both client side and server side. It allows making interactive web pages. It helps in making static HTML web pages into dynamic and interactive one. Web Pages can be made dynamic by dynamically controlling multimedia, updating content, animate images, validating form data.

MYSQL

It is an open source relational database management system. It is developed distributed and supported by Oracle. It works on client/server and also on embedded systems. It consists of multithreaded SQL servers which have support for different client programs, back ends, libraries, administrative tools and a wide range of application programming interfaces (APIs).

DJANGO

It is a high level Python web framework. It encourages maintainable, secure, clean pragmatic design and rapid development of websites. It takes care of most of the hassle of web development, so one can focus on writing an app without having the need to reinvent the wheel.

JUPYTER NOTEBOOK

It is an open source web application. It is used to create and share documents that contain equations, visualizations, live code. It is a client-server application. It provides functionalities to perform all type of data science related tasks data cleaning, data transformation, data visualization, statistical modeling, machine learning, and deep learning. It can be converted into different types of standard output formats using web interface. It provides flexibility and ease in sharing the work with others.

It has two components: Front end web page and a Back end kernel. Front End helps in writing the code and text in the rectangular cells provided. Back end kernel runs this code which is passed by the browser and returns the result.

PYCHARM

It is hybrid platform and an integrated development environment (IDE) for python which is used in computer programming. It provides packages, modules and tools to speed up the Python development. It can be customized according to the requirements and personal preferences.

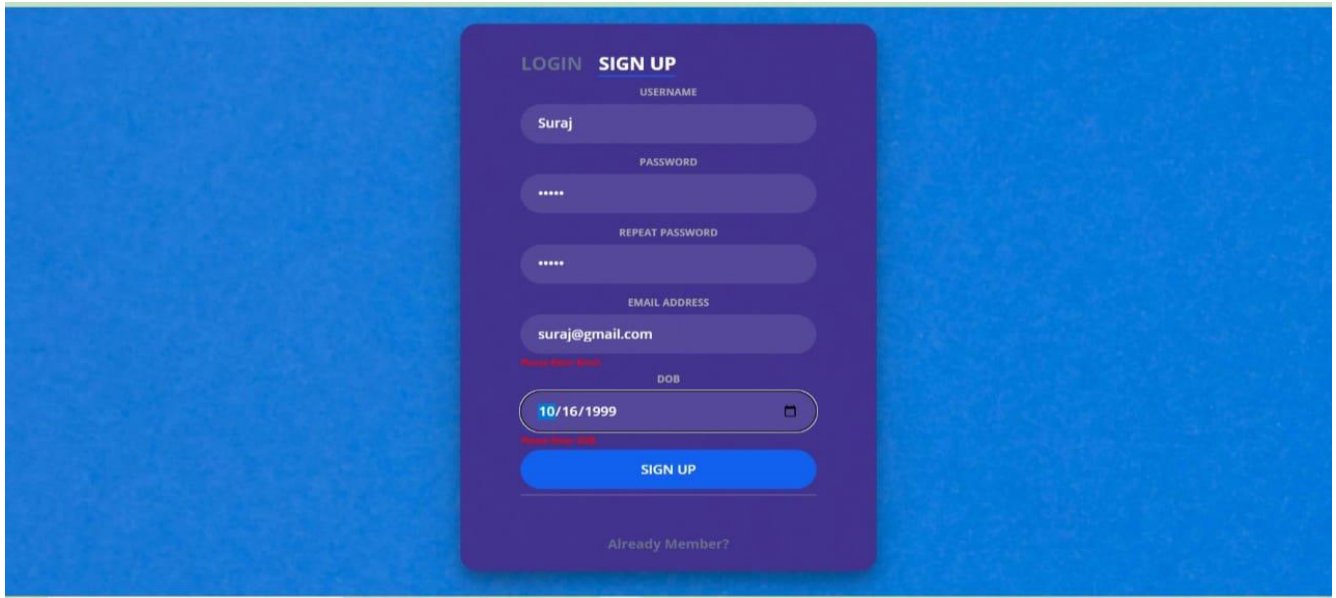
8. EXPERIMENTAL RESULT

	internal_it	toxic_look	depressor	irritability	muscle_pa	altered_se	red_spots	belly_pain	abnormal	dischrimic	watering	family_his	mucoid_s	rusty_sput	lack_of_cc	visual_dist	blood_in	palpitation	painful_wi	pus_filled	blackhead	scurring	prognosis
2	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
4	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
5	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
7	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
9	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
11	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infecti
12	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
13	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
14	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
15	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
19	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
21	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
32	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d
33	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d
34	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d
35	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d
36	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d
37	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer d

Fig 8.1: Training Data

	CF	CG	CH	CI	CJ	CK	CL	CM	CN	CO	CP	CQ	CR	CS	CT	CU
2	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Fungal infection
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0 Allergy
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Drug Reaction
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Peptic ulcer disease
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Gastroenteritis
7	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0 Bronchial Asthma
8	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0 Hypertension
9	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0 Migraine
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Cervical spondylosis
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Paralysis (brain hem
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Jaundice
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Malaria
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Chicken pox
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Dengue
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Typhoid
17	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0 Tuberculosis
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Common Cold
19	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0 Pneumonia
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Dimorphic hemmor
21	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Hypothyroidism
22	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Hyperthyroidism
23	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0 Hypoglycemia
24	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0 Arthritis
25	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1 Acne
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0 Urinary tract infecti

Fig 8.2: Testing Data



The registration page features a dark purple card on a blue background. At the top, 'LOGIN' and 'SIGN UP' are displayed, with 'SIGN UP' being the active tab. The form includes fields for USERNAME (containing 'Suraj'), PASSWORD (masked with '*****'), REPEAT PASSWORD (masked with '*****'), EMAIL ADDRESS (containing 'suraj@gmail.com'), and DOB (containing '10/16/1999' with a calendar icon). Red error messages 'Please enter Email' and 'Please enter DOB' are visible below the email and DOB fields respectively. A blue 'SIGN UP' button is at the bottom of the form, and a link 'Already Member?' is located below it.

LOGIN SIGN UP

USERNAME

Suraj

PASSWORD

REPEAT PASSWORD

EMAIL ADDRESS

suraj@gmail.com

Please enter Email

DOB

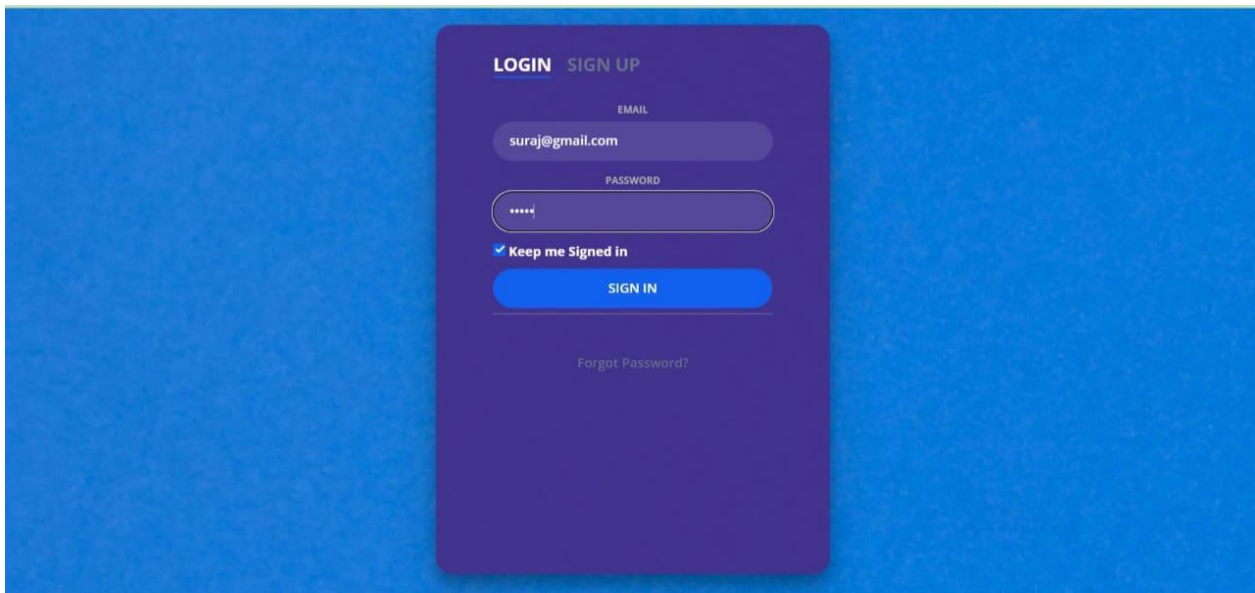
10/16/1999

Please enter DOB

SIGN UP

Already Member?

Fig 8.3: Registration Page



The login page features a dark purple card on a blue background. At the top, 'LOGIN' and 'SIGN UP' are displayed, with 'LOGIN' being the active tab. The form includes fields for EMAIL (containing 'suraj@gmail.com') and PASSWORD (masked with '*****'). A checkbox labeled 'Keep me Signed in' is checked. A blue 'SIGN IN' button is at the bottom of the form, and a link 'Forgot Password?' is located below it.

LOGIN SIGN UP

EMAIL

suraj@gmail.com

PASSWORD

☒ Keep me Signed in

SIGN IN

Forgot Password?

Fig 8.4: Login Page

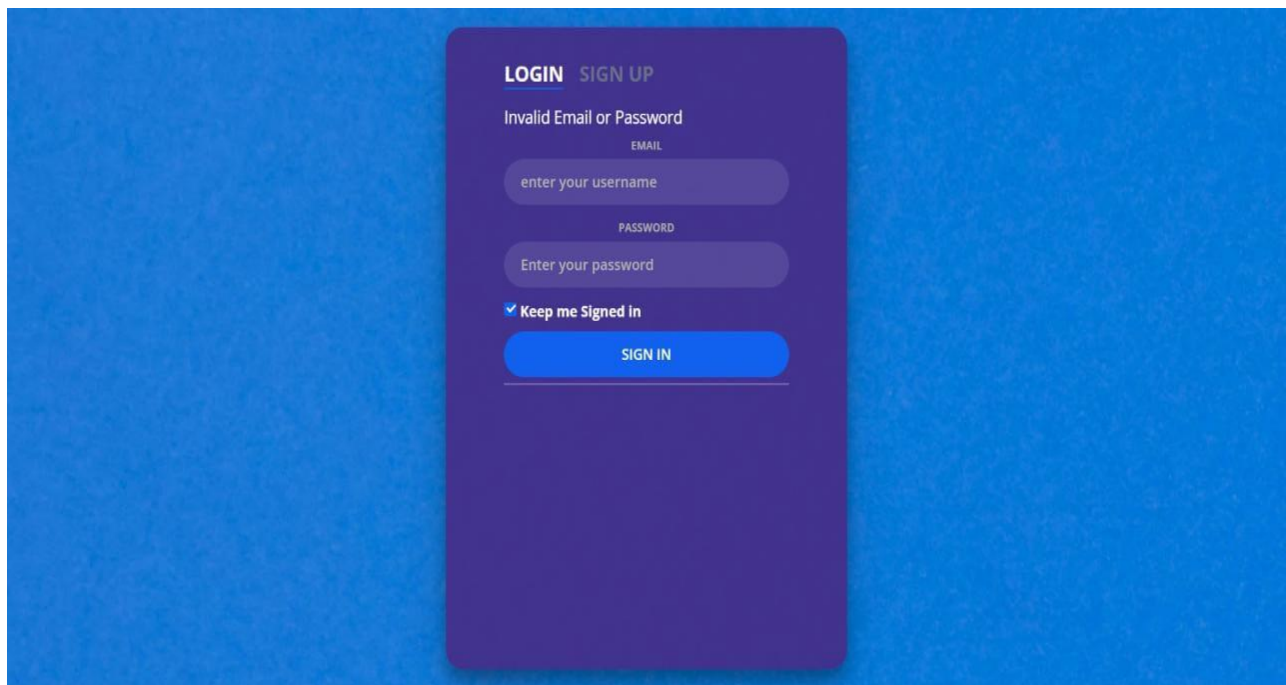


Fig 8.5 Login page showing Invalid Email or Password



Fig 8.6: Dashboard



Fig 8.7: Dashboard with Chosen Symptoms



Fig 8.8: Dashboard Predicting Disease

GDPS
Previous Disease
Contact

Past Disease

Symptoms 1	Symptoms 2	Symptoms 3	Symptoms 4	Symptoms 5	Disease	Date
VISUAL_DISTURBANCES	EXCESSIVE_HUNGER	DEPRESSION	IRRITABILITY	STIFF_NECK	MIGRAINE	2/5/2022
STIFF_NECK	BACK_PAIN	MOOD_SWINGS			ARTHRITIS	2/5/2022
ITCHING	BURNING_MICTURITION	COUGH			DRUG REACTION	2/5/2022
SWELLED_LYMPH_NODES	MILD_FEVER	LOSS_OF_APPETITE	RED_SPOTS_OVER_BODY		CHICKEN POX	2/5/2022


Contact Us

8604105752

Fig 8.9: Past Record of Patient

9. CONCLUSION AND FUTURE ENHANCEMENT

9.1 CONCLUSION

This project “General Disease Prediction using ML” provide prediction for various diseases that are occurring generally in a person and mostly people ignore this which sometimes turn into fatal disease and create a lot of problems for the patient and family also. In the present time, the internet is emerging every day and people are always very much eager to use different new technologies. People mostly prefer to refer to the internet if any problem arises. Nowadays, People have very quick access to the internet than hospitals and doctors as the internet is available all the time and can be accessed from anywhere. Sometimes, people do not have immediate options for doing anything and going to the hospital when they suffer from a health-related issue. For those, this application can be very helpful and they have access to this system with the help of the internet any time in a day.

In conclusion, this project “General Disease Prediction using Machine Learning” will be very much helpful in everyone's daily life. Nowadays the health industry plays a major role in curing the diseases of the patients so this application will provide some help to the health sectors and also it will be useful for the user in case they don't want to go to the hospital or any other clinic. So, just by entering symptoms, the user will get to know about the disease they are suffering from. The work of the doctor can also be reduced, if the health industry adopts this project as it can easily predict the disease of the patient.

9.2 FUTURE ENHANCEMENT

- More Interactive user interface
- Suggestions, Precautions to take based on the predicted disease
- We can add doctor module from where the patient can get the information of the doctor and can directly contact him for medication.
- More details and latest diseases can be inserted.
- It can be converted into Mobile Application.

10. REFERENCES

- [1] Pahulpreet Singh Kohli, Shriya Arora, “APPLICATION OF MACHINE LEARNING IN DISEASE PREDICTION”, 4th International Conference on Computing Communication and Automation (ICCCA), IEEE, 2018.
- [2] Dhiraj Dahiwade, Prof. Gajanan Patle, Prof Ektaa Meshram, “DESIGNING DISEASE PREDICTION MODEL USING MACHINE LEARNING APPROACH”, Proceedings of the Third International Conference on Computing Methodologies and Communication (ICCMC 2019), IEEE, 2019.
- [3] Nikita Pawar, Pooja Sonawane, Narendra Gawai “DECISION TREE BASED HEALTH PREDICTION SYSTEM”, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 8 Issue VII ISSN: 2321-9653, 2020.
- [4] Nishant Yede , Ritik Koul , Chetan Harde , Kumar Gaurav , Prof. C.S.Pagar, “GENERAL DISEASE PREDICTION BASED ON SYMPTOMS PROVIDED BY PATIENT” Open Access International Journal of Science and Engineering, Volume-6, ISSN: 2456-3293, 2021.
- DataSet:-** <https://www.kaggle.com/kaushil268/disease-prediction-using-machine-learning>