NEIL GOGTE INSTITUTE OF TECHNOLOGY

A Unit of Keshav Memorial Technical Education (KMTES)

Approved by AICTE, New Delhi & Affiliated to Osmania University, Hyderabad

A

MINI PROJECT REPORT

ON

DISEASE PROGNOSIS USING RANDOM FOREST

For the award of Degree of

BACHELOR OF ENGINEERING

IN

CSE (AI ML)

Submitted By

D. RAHUL	**245321748022**

Under the guidance of

Mr. P. Nageswara Rao
Assistant Professor

Department of CSE(AIML)

NEIL GOGTE INSTITUTE OF TECHNOLOGY

Kachavanisingaram Village, Hyderabad, Telangana 500058.

FEBRUARY 2024

CERTIFICATE

This is to certify that the Mini project work entitled "**DISEASE PROGNOSIS USING RANDOM FOREST"** *is a bonafide work carried out by* **D. RAHUL(245321748022)** *of III-year V semester* **Bachelor of Engineering** *in CSE(***AIML)** *by Osmania University, Hyderabad during the academic year* **2023-2024** *is a record of bonafide work carried out by them. The results embodied in this report have not been submitted to any other University or Institution for the award of any degree*

Internal Guide   **Head of Department**

Mr. P. Nageswara Rao

Assistant Professor

Dr. T. Prem Chander

Associate Professor

External

DECLARATION

I hereby declare that the Mini Project Report entitled, "**DIESEASE PROGNOSIS USING Random Forest**" submitted for the B.E degree is entirely my work and all ideas and references have been duly acknowledged. It does not contain any work for the award of any other degree.

Date:

D. RAHUL   **245321748022**

ACKNOWLEDGEMENT

ABSTRACT

Technological Landscape: In the contemporary landscape dominated by Machine Learning and Deep Learning, there exists a critical need to bridge the gap between theoretical knowledge and its practical application, especially in fields crucial to human welfare.

Purpose of the Project: This project is conceived with the explicit goal of channeling the burgeoning expertise in machine learning towards the realm of Medical Science. The overarching objective is to simplify the complexities of medical diagnosis by providing a dataset designed for predictive analysis.

Dataset Overview: A dataset comprising 132 parameters has been meticulously curated, presenting a unique opportunity for practitioners to harness their existing knowledge in machine learning. The dataset enables the prediction of 42 distinct diseases, covering a diverse range of medical conditions.

TABLE OF CONTENTS

List of Figures

List of Tables

V

CHAPTER – 1
INTRODUCTION

1.1 PROBLEM STATEMENT

Accurate diagnosis of diseases based on symptoms can be challenging due to the complexity and variability of human health conditions. Traditional diagnostic methods may rely on subjective interpretation or limited medical knowledge, leading to misdiagnosis or delayed treatment.

1.2 MOTIVATION

The motivation behind this project lies in the potential to revolutionize disease prognosis by leveraging machine learning techniques. By developing a model capable of accurately predicting diseases based on symptoms, we aim to enhance healthcare delivery, improve patient outcomes, and optimize resource allocation in the medical field.

1.3 SCOPE

This project focuses on developing a machine learning-based system for health prognosis using symptoms provided by the user. The scope encompasses a wide range of diseases, with a primary focus on common ailments such as respiratory infections, gastrointestinal disorders, and cardiovascular diseases. The system will consider a variety of symptoms, including but not limited to fever, cough, chest pain, nausea, and fatigue. However, it is important to note that the system's performance may vary depending on the complexity and specificity of the disease.

1.4 OUTLINE

The VGG16 CNN model classifies the dataset into three parts, i.e., training dataset, testing dataset and lastly, validation dataset. The model learns from the training dataset and then using the validation dataset, inputs are validated. I used "Streamlit" to create web application that is flexible to use 200MB sized images as its input.

CHAPTER – 2

LITERATURE SURVEY

EXISTING SYSTEM:

In the domain of health prognosis using machine learning (ML), prior research has emphasized the importance of feature extraction or classification from symptom data to improve diagnostic accuracy. The effectiveness of these systems is heavily dependent on the quality and relevance of the features extracted from the input symptoms. Within the medical domain, researchers have explored various methodologies to differentiate between different health conditions based on symptom patterns. For example, a study introduced a novel feature selection approach leveraging statistical methods and domain knowledge to identify key symptoms associated with specific diseases. These features were then integrated into machine learning models such as random forests and decision trees, achieving notable success in accurately predicting disease prognosis from symptom data.

PROPOSED SYSTEM:

The proposed system employs a Random Forest algorithm, a popular machine learning ensemble method, for health prognosis based on user-provided symptoms. Unlike traditional machine learning algorithms, Random Forest requires minimal preprocessing of symptom data and can effectively handle both numerical and categorical features. The algorithm operates by constructing multiple decision trees during training and combining their predictions through voting to determine the final prognosis. This ensemble approach enhances the robustness and generalization of the model, making it suitable for handling diverse symptom patterns and predicting various health conditions. The proposed system aims to leverage Random Forest for disease prognosis by utilizing symptom data provided by users. By analyzing the symptom profiles and learning from historical datasets, the Random Forest model can effectively identify patterns and associations indicative of different diseases. Through rigorous training and validation procedures, the model is expected to achieve high accuracy and reliability in predicting disease prognosis based on user symptoms.

CHAPTER - 3

SOFTWARE REQUIREMENTS SPECIFICATION

3.1 Overall Description:

This SRS is an overview of the whole project scenario. This document is to present a detailed description of the course management system. It will explain the purpose and features of the system, the interfaces of the system will do, the constraints under which it must operate and how the system will react to external stimuli. This document is intended for both stakeholders and developers of the system.

3.2. Operating Environment:

Software Requirements:

Operating System	:	Windows 7 (Min)
Front End	:	Html, CSS, JS
Back End	:	Nodejs ,Express
Database	:	Microsoft Excel

Hardware Requirements:

Processor	:	Intel Pentium® Dual Core Processor (Min)
Speed	:	2.9 GHz (Min)
RAM	:	2 GB (Min)
Hard Disk	:	2 GB (Min)

3.3 Functional Requirements:

User Functionality:

- The user will be able to select symptoms regarding the disease he/she wants to find
- The user will be able to insert symptoms of 132 types.
- The user can see information regarding the disease he/she will also be provided with disease's Wikipedia.

Admin Functionality:

- The admin manages the website.

- The admin can increase database size.

- The admin can make changes to the website such as modifying the UI and making it more interactive than earlier.
- The admin can implement a better algorithm if at all a better algorithm is created in future.

3.4 Non-Functional Requirements:

3.4.1 Performance Requirements:

Performance requirements refer to static numerical requirements placed on the interaction betIen the users and the software.

Response Time:

Average response time shall be less than 5 sec.

Recovery Time:

In case of system failure, the redundant system shall resume operations within 30 secs. Average repair time shall be less than 45 minutes.

Start-Up/Shutdown Time:

The system shall be operational within 1 minute of starting up.

Capacity:

The system accommodates 1000 Concurrent Users.

Utilization of Resources:

The system shall store in the database no more than 450 different species with room for improvement.

3.4.2 **Safety Requirements:**

-NA-

3.4.3 **Security Requirements:**

The model will be running on a secure website i.e., an HTTPS website and on a secure browser such as Google Chrome, Brave, etc.

3.4.4 **Software Quality Attributes:**

Reliability:

The system shall be reliable i.e., in case the webpage crashes, progress will be saved.

Availability:

The website will be available to all its users round the clock i.e., they can access the website at any time.

Security:

The model will be running on a secure website i.e., an HTTPS website and on a secure browser such as Google Chrome, Brave, etc.

Maintainability:

The model shall be designed in such a way that it will be very easy to maintain it in future. Our model is a neural network model and a web-based system and will depend much on the web server and on the neural networks. However, the web application will be designed using Streamlit which is based on neural network approach and proper database modeling along with extensive documentation which will make it easy to develop, troubleshoot and maintain in future.

Usability:

The interfaces of the system will be user friendly enough that every user will be able to use it easily.

Scalability:

The system will be designed in such a way that it will be extendable. If more species or algorithms are going to be added in the system, then it would easily be done.

The same system can also be developed to become a mobile application rather than just a website.

CHAPTER-4

SYSTEM DESIGN

Use case Diagrams:
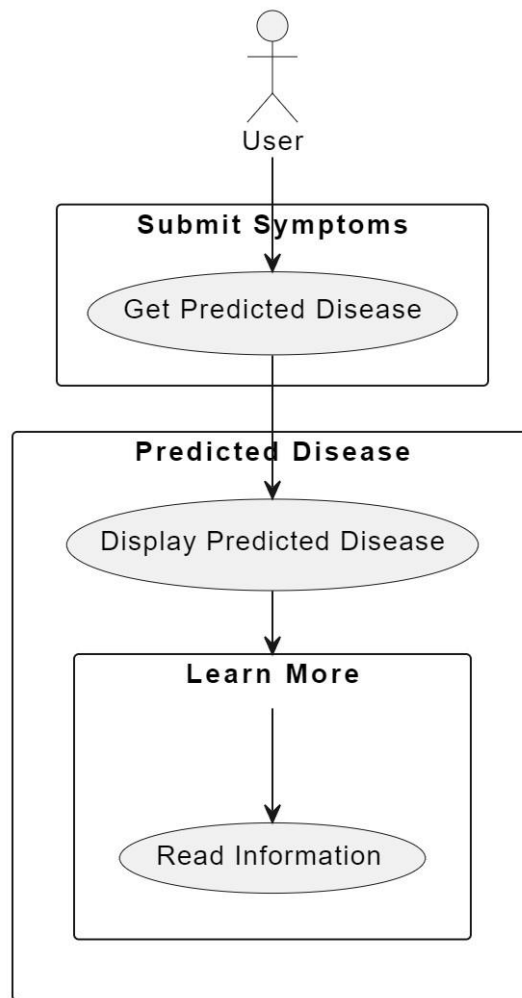
Fig 4.1: Use case diagram for User
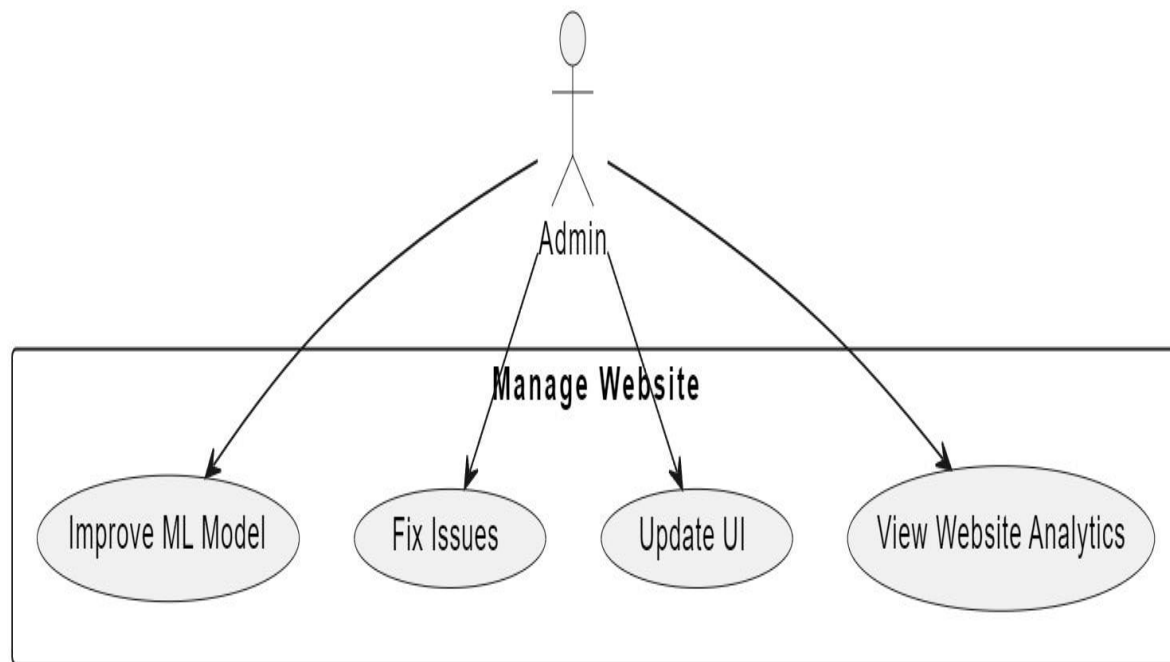
Fig 4.2: Use case diagram for Admin

Class Diagram:
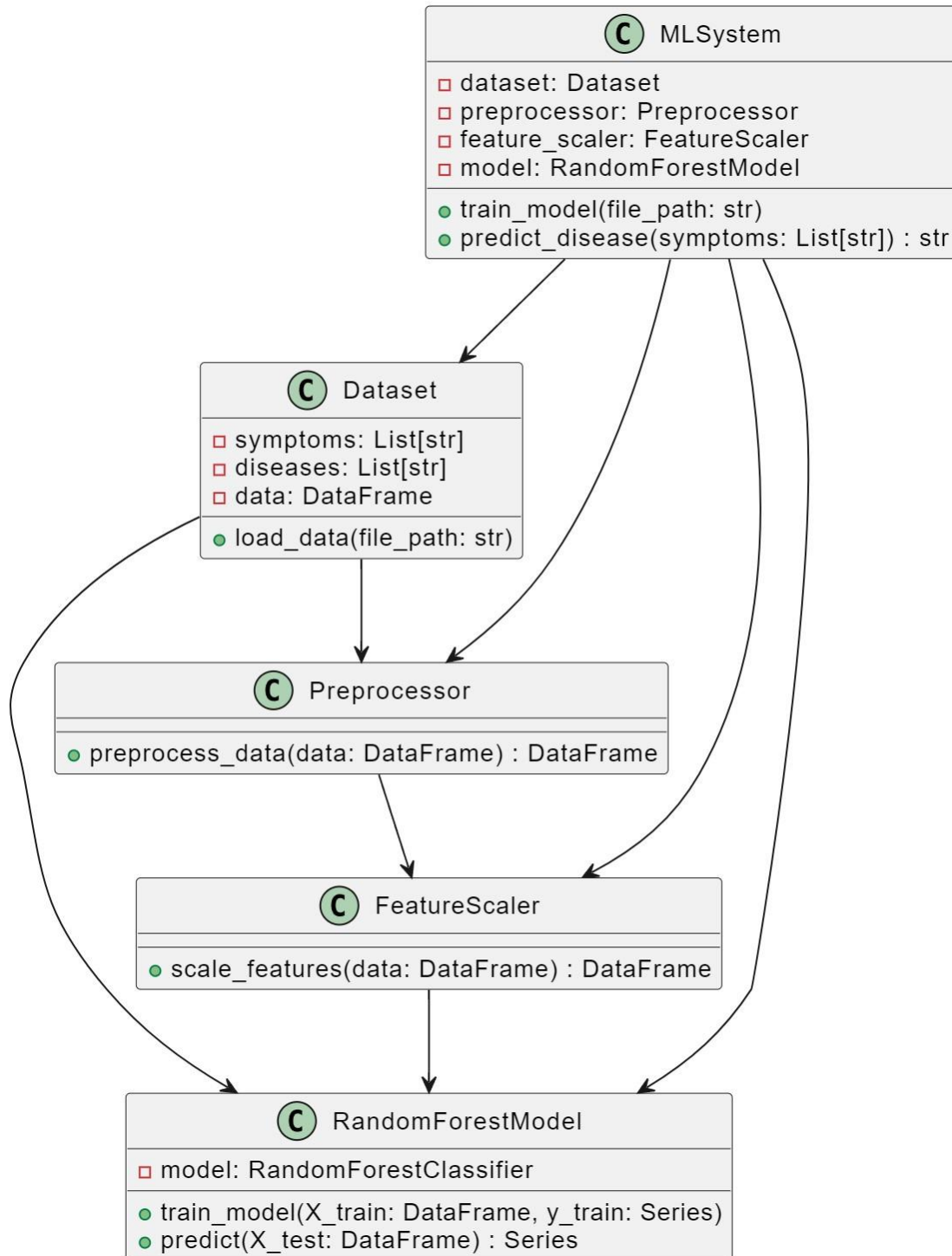
Fig.4.3 : Class diagram for Model Training.

Sequence Diagram:
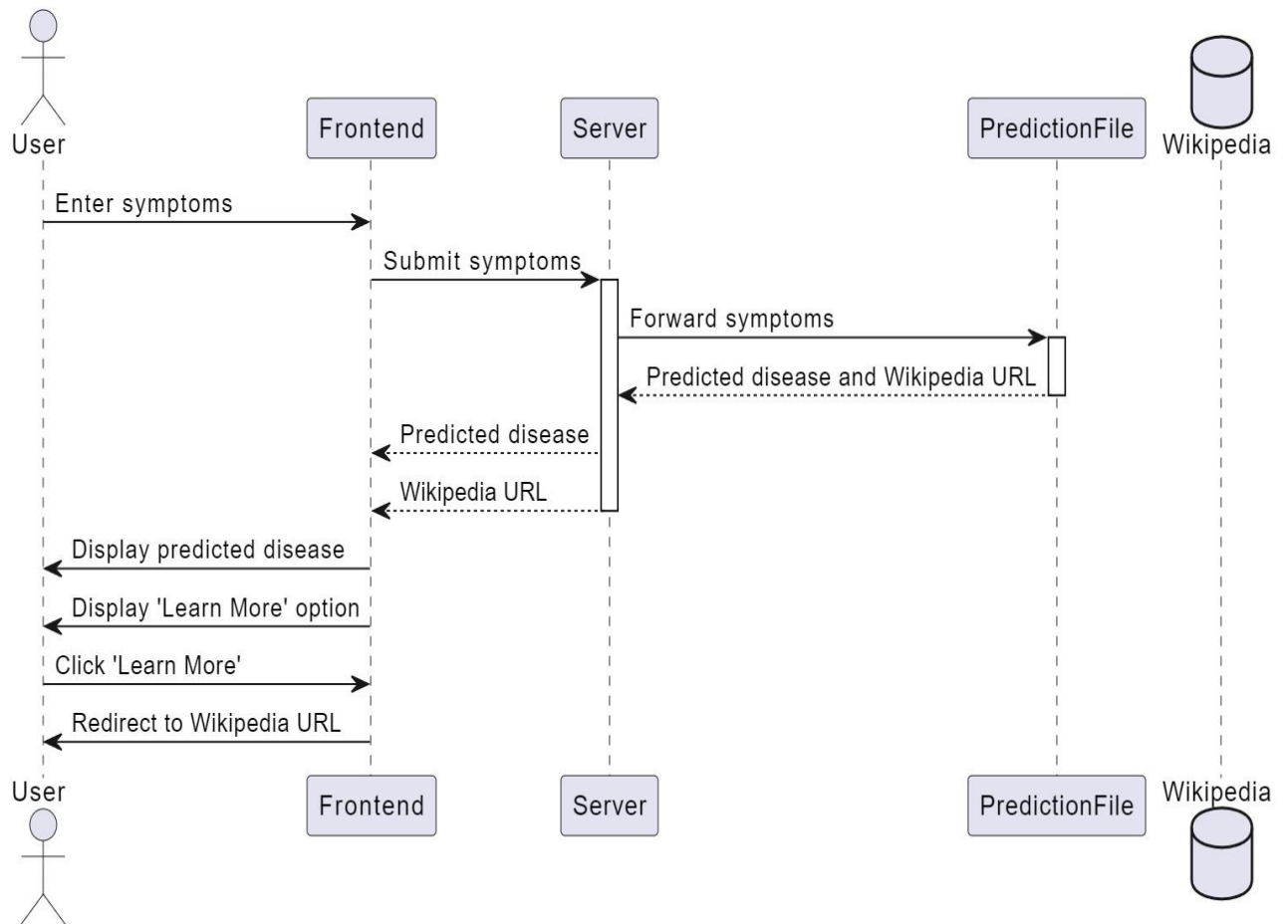
Fig4.3 : Sequence Diagam for Disease Prediction.

CHAPTER – 5

IMPLEMENTATION

5.1 SAMPLE CODE

Model Training Code :-

```python
import pandas as pd
import numpy as np

df = pd.read_csv('Training.csv')
df.info(verbose=True)
df
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919
Data columns (total 134 columns):
 #   Column                   Dtype
---  ------                   -----
 0   itching                  int64
 1   skin_rash                int64
 2   nodal_skin_eruptions     int64
 3   continuous_sneezing      int64
 4   shivering                int64
 5   chills                   int64
 6   joint_pain               int64
 7   stomach_pain             int64
 8   acidity                  int64
 9   ulcers_on_tongue         int64
 10  muscle_wasting           int64
 11  vomiting                 int64
 12  burning_micturition      int64
 13  spotting_ urination      int64
 14  fatigue                  int64
 15  weight_gain              int64
 16  anxiety                  int64
 17  cold_hands_and_feets     int64
 18  mood_swings              int64
 19  weight_loss              int64
 20  restlessness             int64
 21  lethargy                 int64
 22  patches_in_throat        int64
 23  irregular_sugar_level    int64
 24  cough                    int64
 25  high_fever               int64
 26  sunken_eyes              int64
 27  breathlessness           int64
 28  sweating                 int64
 29  dehydration              int64
 30  indigestion              int64
 31  headache                 int64
 32  yellowish_skin           int64
 33  dark_urine               int64
 34  nausea                   int64
 35  loss_of_appetite         int64
```

```
36  pain_behind_the_eyes          int64
37  back_pain                     int64
38  constipation                  int64
39  abdominal_pain                int64
40  diarrhoea                     int64
41  mild_fever                    int64
42  yellow_urine                  int64
43  yellowing_of_eyes             int64
44  acute_liver_failure           int64
45  fluid_overload                int64
46  swelling_of_stomach           int64
47  swelled_lymph_nodes           int64
48  malaise                       int64
49  blurred_and_distorted_vision  int64
50  phlegm                        int64
51  throat_irritation             int64
52  redness_of_eyes               int64
53  sinus_pressure                int64
54  runny_nose                    int64
55  congestion                    int64
56  chest_pain                    int64
57  weakness_in_limbs             int64
58  fast_heart_rate               int64
59  pain_during_bowel_movements   int64
60  pain_in_anal_region           int64
61  bloody_stool                  int64
62  irritation_in_anus            int64
63  neck_pain                     int64
64  dizziness                     int64
65  cramps                        int64
66  bruising                      int64
67  obesity                       int64
68  swollen_legs                  int64
69  swollen_blood_vessels         int64
70  puffy_face_and_eyes           int64
71  enlarged_thyroid              int64
72  brittle_nails                 int64
73  swollen_extremeties           int64
74  excessive_hunger              int64
75  extra_marital_contacts        int64
76  drying_and_tingling_lips      int64
77  slurred_speech                int64
78  knee_pain                     int64
79  hip_joint_pain                int64
80  muscle_weakness               int64
81  stiff_neck                    int64
82  swelling_joints               int64
83  movement_stiffness            int64
84  spinning_movements            int64
85  loss_of_balance               int64
86  unsteadiness                  int64
87  weakness_of_one_body_side     int64
88  loss_of_smell                 int64
89  bladder_discomfort            int64
90  foul_smell_of urine           int64
91  continuous_feel_of_urine      int64
92  passage_of_gases              int64
93  internal_itching              int64
94  toxic_look_(typhos)           int64
95  depression                    int64
96  irritability                  int64
```

```
97   muscle_pain                   int64
98   altered_sensorium             int64
99   red_spots_over_body           int64
100  belly_pain                    int64
101  abnormal_menstruation         int64
102  dischromic _patches           int64
103  watering_from_eyes            int64
104  increased_appetite            int64
105  polyuria                      int64
106  family_history                int64
107  mucoid_sputum                 int64
108  rusty_sputum                  int64
109  lack_of_concentration         int64
110  visual_disturbances           int64
111  receiving_blood_transfusion   int64
112  receiving_unsterile_injections int64
113  coma                          int64
114  stomach_bleeding              int64
115  distention_of_abdomen         int64
116  history_of_alcohol_consumption int64
117  fluid_overload.1              int64
118  blood_in_sputum               int64
119  prominent_veins_on_calf       int64
120  palpitations                  int64
121  painful_walking               int64
122  pus_filled_pimples            int64
123  blackheads                    int64
124  scurring                      int64
125  skin_peeling                  int64
126  silver_like_dusting           int64
127  small_dents_in_nails          int64
128  inflammatory_nails            int64
129  blister                       int64
130  red_sore_around_nose          int64
131  yellow_crust_ooze             int64
132  prognosis                     object
133  Unnamed: 133                  float64
dtypes: float64(1), int64(132), object(1)
memory usage: 5.0+ MB
```

	itching	skin_rash	nodal_skin_eruptions	continuous_sneezing
0	1	1	1	0
1	0	1	1	0
2	1	0	1	0
3	1	1	0	0
4	1	1	1	0
...	...	...	...	...
4915	0	0	0	0
4916	0	1	0	0
4917	0	0	0	0
4918	0	1	0	0
4919	0	1	0	0

	shivering	chills	joint_pain	stomach_pain	acidity	ulcers_on_tongue
0	0	0	0	0	0	0
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
...	...	...	...	...	...	...
4915	0	0	0	0	0	0

```
4916      0    0      0        0    0          0
4917      0    0      0        0    0          0
4918      0    0      1        0    0          0
4919      0    0      0        0    0          0

      ... scurring skin_peeling silver_like_dusting small_dents_in_nails
0     ...     0        0            0                0
1     ...     0        0            0                0
2     ...     0        0            0                0
3     ...     0        0            0                0
4     ...     0        0            0                0
...   ...    ...      ...          ...              ...
4915  ...     0        0            0                0
4916  ...     1        0            0                0
4917  ...     0        0            0                0
4918  ...     0        1            1                1
4919  ...     0        0            0                0

      inflammatory_nails blister red_sore_around_nose yellow_crust_ooze
0             0      0          0            0
1             0      0          0            0
2             0      0          0            0
3             0      0          0            0
4             0      0          0            0
...          ...    ...        ...          ...
4915          0      0          0            0
4916          0      0          0            0
4917          0      0          0            0
4918          1      0          0            0
4919          0      1          1            1

                     prognosis  Unnamed: 133
0             Fungal infection      NaN
1             Fungal infection      NaN
2             Fungal infection      NaN
3             Fungal infection      NaN
4             Fungal infection      NaN
...                  ...          ...
4915  (vertigo) Paroymsal  Positional Vertigo      NaN
4916                    Acne      NaN
4917        Urinary tract infection          NaN
4918                 Psoriasis      NaN
4919                  Impetigo      NaN

[4920 rows x 134 columns]
```

print(df.columns.tolist())

['itching', 'skin_rash', 'nodal_skin_eruptions', 'continuous_sneezing', 'shivering', 'chills', 'joint_pain', 'stomach_pain', 'acidity', 'ulcers_on_tongue', 'muscle_wasting', 'vomiting', 'burning_micturition', 'spotting_ urination', 'fatigue', 'weight_gain', 'anxiety', 'cold_hands_and_feets', 'mood_swings', 'weight_loss', 'restlessness', 'lethargy', 'patches_in_throat', 'irregular_sugar_level', 'cough', 'high_fever', 'sunken_eyes', 'breathlessness', 'sweating', 'dehydration', 'indigestion', 'headache', 'yellowish_skin', 'dark_urine', 'nausea', 'loss_of_appetite', 'pain_behind_the_eyes', 'back_pain', 'constipation', 'abdominal_pain', 'diarrhoea', 'mild_fever', 'yellow_urine', 'yellowing_of_eyes', 'acute_liver_failure', 'fluid_overload', 'swelling_of_stomach', 'swelled_lymph_nodes', 'malaise',

'blurred_and_distorted_vision', 'phlegm', 'throat_irritation', 'redness_of_eyes', 'sinus_pressure',

'runny_nose', 'congestion', 'chest_pain', 'weakness_in_limbs', 'fast_heart_rate',

'pain_during_bowel_movements', 'pain_in_anal_region', 'bloody_stool', 'irritation_in_anus', 'neck_pain',

'dizziness', 'cramps', 'bruising', 'obesity', 'swollen_legs', 'swollen_blood_vessels', 'puffy_face_and_eyes',

'enlarged_thyroid', 'brittle_nails', 'swollen_extremeties', 'excessive_hunger', 'extra_marital_contacts',

'drying_and_tingling_lips', 'slurred_speech', 'knee_pain', 'hip_joint_pain', 'muscle_weakness',

'stiff_neck', 'swelling_joints', 'movement_stiffness', 'spinning_movements', 'loss_of_balance',

'unsteadiness', 'weakness_of_one_body_side', 'loss_of_smell', 'bladder_discomfort', 'foul_smell_of

urine', 'continuous_feel_of_urine', 'passage_of_gases', 'internal_itching', 'toxic_look_(typhos)',

'depression', 'irritability', 'muscle_pain', 'altered_sensorium', 'red_spots_over_body', 'belly_pain',

'abnormal_menstruation', 'dischromic _patches', 'watering_from_eyes', 'increased_appetite', 'polyuria',

'family_history', 'mucoid_sputum', 'rusty_sputum', 'lack_of_concentration', 'visual_disturbances',

'receiving_blood_transfusion', 'receiving_unsterile_injections', 'coma', 'stomach_bleeding',

'distention_of_abdomen', 'history_of_alcohol_consumption', 'fluid_overload.1', 'blood_in_sputum',

'prominent_veins_on_calf', 'palpitations', 'painful_walking', 'pus_filled_pimples', 'blackheads', 'scurring',

'skin_peeling', 'silver_like_dusting', 'small_dents_in_nails', 'inflammatory_nails', 'blister',

'red_sore_around_nose', 'yellow_crust_ooze', 'prognosis', 'Unnamed: 133']

```
df.drop('Unnamed: 133',axis=1,inplace=True)

df.info(verbose=True, show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4920 entries, 0 to 4919
Data columns (total 133 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   itching               4920 non-null   int64
 1   skin_rash             4920 non-null   int64
 2   nodal_skin_eruptions  4920 non-null   int64
 3   continuous_sneezing   4920 non-null   int64
 4   shivering             4920 non-null   int64
 5   chills                4920 non-null   int64
 6   joint_pain            4920 non-null   int64
 7   stomach_pain          4920 non-null   int64
 8   acidity               4920 non-null   int64
 9   ulcers_on_tongue      4920 non-null   int64
 10  muscle_wasting        4920 non-null   int64
 11  vomiting              4920 non-null   int64
 12  burning_micturition   4920 non-null   int64
 13  spotting_ urination   4920 non-null   int64
 14  fatigue               4920 non-null   int64
 15  weight_gain           4920 non-null   int64
 16  anxiety               4920 non-null   int64
 17  cold_hands_and_feets  4920 non-null   int64
 18  mood_swings           4920 non-null   int64
 19  weight_loss           4920 non-null   int64
 20  restlessness          4920 non-null   int64
```

```
21  lethargy                    4920 non-null   int64
22  patches_in_throat           4920 non-null   int64
23  irregular_sugar_level       4920 non-null   int64
24  cough                       4920 non-null   int64
25  high_fever                  4920 non-null   int64
26  sunken_eyes                 4920 non-null   int64
27  breathlessness              4920 non-null   int64
28  sweating                    4920 non-null   int64
29  dehydration                 4920 non-null   int64
30  indigestion                 4920 non-null   int64
31  headache                    4920 non-null   int64
32  yellowish_skin              4920 non-null   int64
33  dark_urine                  4920 non-null   int64
34  nausea                      4920 non-null   int64
35  loss_of_appetite            4920 non-null   int64
36  pain_behind_the_eyes        4920 non-null   int64
37  back_pain                   4920 non-null   int64
38  constipation                4920 non-null   int64
39  abdominal_pain              4920 non-null   int64
40  diarrhoea                   4920 non-null   int64
41  mild_fever                  4920 non-null   int64
42  yellow_urine                4920 non-null   int64
43  yellowing_of_eyes           4920 non-null   int64
44  acute_liver_failure         4920 non-null   int64
45  fluid_overload              4920 non-null   int64
46  swelling_of_stomach         4920 non-null   int64
47  swelled_lymph_nodes         4920 non-null   int64
48  malaise                     4920 non-null   int64
49  blurred_and_distorted_vision  4920 non-null   int64
50  phlegm                      4920 non-null   int64
51  throat_irritation           4920 non-null   int64
52  redness_of_eyes             4920 non-null   int64
53  sinus_pressure              4920 non-null   int64
54  runny_nose                  4920 non-null   int64
55  congestion                  4920 non-null   int64
56  chest_pain                  4920 non-null   int64
57  weakness_in_limbs           4920 non-null   int64
58  fast_heart_rate             4920 non-null   int64
59  pain_during_bowel_movements  4920 non-null   int64
60  pain_in_anal_region         4920 non-null   int64
61  bloody_stool                4920 non-null   int64
62  irritation_in_anus          4920 non-null   int64
63  neck_pain                   4920 non-null   int64
64  dizziness                   4920 non-null   int64
65  cramps                      4920 non-null   int64
66  bruising                    4920 non-null   int64
67  obesity                     4920 non-null   int64
68  swollen_legs                4920 non-null   int64
69  swollen_blood_vessels       4920 non-null   int64
70  puffy_face_and_eyes         4920 non-null   int64
71  enlarged_thyroid            4920 non-null   int64
72  brittle_nails               4920 non-null   int64
73  swollen_extremeties         4920 non-null   int64
74  excessive_hunger            4920 non-null   int64
75  extra_marital_contacts      4920 non-null   int64
76  drying_and_tingling_lips    4920 non-null   int64
77  slurred_speech              4920 non-null   int64
78  knee_pain                   4920 non-null   int64
79  hip_joint_pain              4920 non-null   int64
80  muscle_weakness             4920 non-null   int64
81  stiff_neck                  4920 non-null   int64
```

```
82  swelling_joints            4920 non-null  int64
83  movement_stiffness         4920 non-null  int64
84  spinning_movements         4920 non-null  int64
85  loss_of_balance            4920 non-null  int64
86  unsteadiness               4920 non-null  int64
87  weakness_of_one_body_side  4920 non-null  int64
88  loss_of_smell              4920 non-null  int64
89  bladder_discomfort         4920 non-null  int64
90  foul_smell_of urine        4920 non-null  int64
91  continuous_feel_of_urine   4920 non-null  int64
92  passage_of_gases           4920 non-null  int64
93  internal_itching           4920 non-null  int64
94  toxic_look_(typhos)        4920 non-null  int64
95  depression                 4920 non-null  int64
96  irritability               4920 non-null  int64
97  muscle_pain                4920 non-null  int64
98  altered_sensorium          4920 non-null  int64
99  red_spots_over_body        4920 non-null  int64
100 belly_pain                 4920 non-null  int64
101 abnormal_menstruation      4920 non-null  int64
102 dischromic _patches        4920 non-null  int64
103 watering_from_eyes         4920 non-null  int64
104 increased_appetite         4920 non-null  int64
105 polyuria                   4920 non-null  int64
106 family_history             4920 non-null  int64
107 mucoid_sputum              4920 non-null  int64
108 rusty_sputum               4920 non-null  int64
109 lack_of_concentration      4920 non-null  int64
110 visual_disturbances        4920 non-null  int64
111 receiving_blood_transfusion 4920 non-null  int64
112 receiving_unsterile_injections 4920 non-null  int64
113 coma                       4920 non-null  int64
114 stomach_bleeding           4920 non-null  int64
115 distention_of_abdomen      4920 non-null  int64
116 history_of_alcohol_consumption 4920 non-null  int64
117 fluid_overload.1           4920 non-null  int64
118 blood_in_sputum            4920 non-null  int64
119 prominent_veins_on_calf    4920 non-null  int64
120 palpitations               4920 non-null  int64
121 painful_walking            4920 non-null  int64
122 pus_filled_pimples         4920 non-null  int64
123 blackheads                 4920 non-null  int64
124 scurring                   4920 non-null  int64
125 skin_peeling               4920 non-null  int64
126 silver_like_dusting        4920 non-null  int64
127 small_dents_in_nails       4920 non-null  int64
128 inflammatory_nails         4920 non-null  int64
129 blister                    4920 non-null  int64
130 red_sore_around_nose       4920 non-null  int64
131 yellow_crust_ooze          4920 non-null  int64
132 prognosis                  4920 non-null  object
dtypes: int64(132), object(1)
memory usage: 5.0+ MB
```

```python
from sklearn.preprocessing import LabelEncoder
LE = LabelEncoder()
df['prognosis'] = LE.fit_transform(df['prognosis'])
df['prognosis']
```

```
0     15
1     15
```

```
2      15
3      15
4      15
       ..
4915    0
4916    2
4917   38
4918   35
4919   27
Name: prognosis, Length: 4920, dtype: int32
```

```python
from sklearn.model_selection import train_test_split
X = df.iloc[:,:132]
y = df['prognosis']
```

```python
from sklearn.preprocessing import StandardScaler
std = StandardScaler()
X = std.fit_transform(X)
```

```python
X_train,X_test,y_train,y_test = train_test_split(X,y,test_size=0.3,random_state=0)
print(y_train.info())
print(y_test.info())
```

```
<class 'pandas.core.series.Series'>
Int64Index: 3444 entries, 3972 to 2732
Series name: prognosis
Non-Null Count  Dtype
--------------  -----
3444 non-null   int32
dtypes: int32(1)
memory usage: 40.4 KB
None
<class 'pandas.core.series.Series'>
Int64Index: 1476 entries, 1525 to 3275
Series name: prognosis
Non-Null Count  Dtype
--------------  -----
1476 non-null   int32
dtypes: int32(1)
memory usage: 17.3 KB
None
```

```python
from sklearn.ensemble import RandomForestClassifier
RFC = RandomForestClassifier()
RFC.fit(X_train,y_train)
```

```
RandomForestClassifier()
```

```python
y_pred = RFC.predict(X_test)
y_pred
```

```
array([18, 40, 36, ..., 30, 39,  0])
```

```python
from sklearn.metrics import accuracy_score
print(accuracy_score(y_test,y_pred)*100)
```

```
100.0
```

```python
y_train_pred = RFC.predict(X_train)
```

```python
print(accuracy_score(y_train,y_train_pred)*100)
```

100.0

```python
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.ensemble import RandomForestClassifier

model = RandomForestClassifier(n_estimators=100)


num_folds = 5

kfold = KFold(n_splits=num_folds, shuffle=True, random_state=42)

results = cross_val_score(model, X, y, cv=kfold)

print("Cross-Validation Results:")
print("Accuracy: %.2f%%" % (results.mean() * 100))
print("Standard Deviation: %.2f%%" % (results.std() * 100))
```

Cross-Validation Results:
Accuracy: 100.00%
Standard Deviation: 0.00%

```python
from sklearn.metrics import classification_report
cr = classification_report(y_test,y_pred)
print(cr)
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	32
1	1.00	1.00	1.00	31
2	1.00	1.00	1.00	38
3	1.00	1.00	1.00	34
4	1.00	1.00	1.00	33
5	1.00	1.00	1.00	36
6	1.00	1.00	1.00	39
7	1.00	1.00	1.00	41
8	1.00	1.00	1.00	33
9	1.00	1.00	1.00	35
10	1.00	1.00	1.00	42
11	1.00	1.00	1.00	27
12	1.00	1.00	1.00	33
13	1.00	1.00	1.00	40
14	1.00	1.00	1.00	40
15	1.00	1.00	1.00	31
16	1.00	1.00	1.00	42
17	1.00	1.00	1.00	36
18	1.00	1.00	1.00	49
19	1.00	1.00	1.00	31
20	1.00	1.00	1.00	32
21	1.00	1.00	1.00	41
22	1.00	1.00	1.00	37
23	1.00	1.00	1.00	35
24	1.00	1.00	1.00	34
25	1.00	1.00	1.00	21
26	1.00	1.00	1.00	33
27	1.00	1.00	1.00	39
28	1.00	1.00	1.00	36
29	1.00	1.00	1.00	31
30	1.00	1.00	1.00	42

31	1.00	1.00	1.00	37
32	1.00	1.00	1.00	41
33	1.00	1.00	1.00	38
34	1.00	1.00	1.00	34
35	1.00	1.00	1.00	31
36	1.00	1.00	1.00	37
37	1.00	1.00	1.00	42
38	1.00	1.00	1.00	36
39	1.00	1.00	1.00	38
40	1.00	1.00	1.00	38
accuracy			1.00	1476
macro avg	1.00	1.00	1.00	1476
weighted avg	1.00	1.00	1.00	1476

CHAPTER – 6

TESTING

6.1 TEST CASES

Test Case ID:	1
Test Case Name:	Required Software Testing
Purpose:	To check whether the required Software is installed on the systems
Input:	Enter python command
Expected Result:	Should Display the version number for the python
Actual Result:	Displays python version
Failure	If the python environment is not installed, then the Deployment fails

Test Case to check whether the required Software is installed on the systems

Table 6.1.1 python Installation verification

Test Case to check Program Integration Testing

Test Case ID:	2
Test Case Name:	Programs Integration Testing
Purpose:	To ensure that all the modules work together
Input:	All the modules should be accessed.
Expected Result:	All the modules should be functioning properly.
Actual Result:	All the modules should be functioning properly.
Failure	If any module fails to function properly, the implementation fails.

Table 6.1.2 python Programs Integration Testing

Test Case to Collect Dataset and Load the Dataset

Test Case ID:	3
Test Case Name:	Collect Dataset and Load the Dataset
Purpose:	Check Dataset is collected, and the data is stored
Input:	Provide Dataset as input
Expected Result:	Dataset is collected and view the Dataset and store the Dataset
Actual Result:	Load the Dataset and view the Dataset and store
Failure	If the dataset is not loaded, it will throw an error.

Table 6.1.3 Collect Dataset and Load the Dataset

Test Case to check whether the species is recognized

Test Case ID:	4
Test Case Name:	Disease Prediction
Purpose:	Disease Prediction using RandomForestClassifier
Input:	Provide dataset and input the symptoms
Expected Result:	After Evaluation I get the Disease Predicted
Actual Result:	After Evaluation I get the Disease Predicted
Failure	If the data is not Evaluated, it does not show anything

Table 6.1.4 Disease Prediction

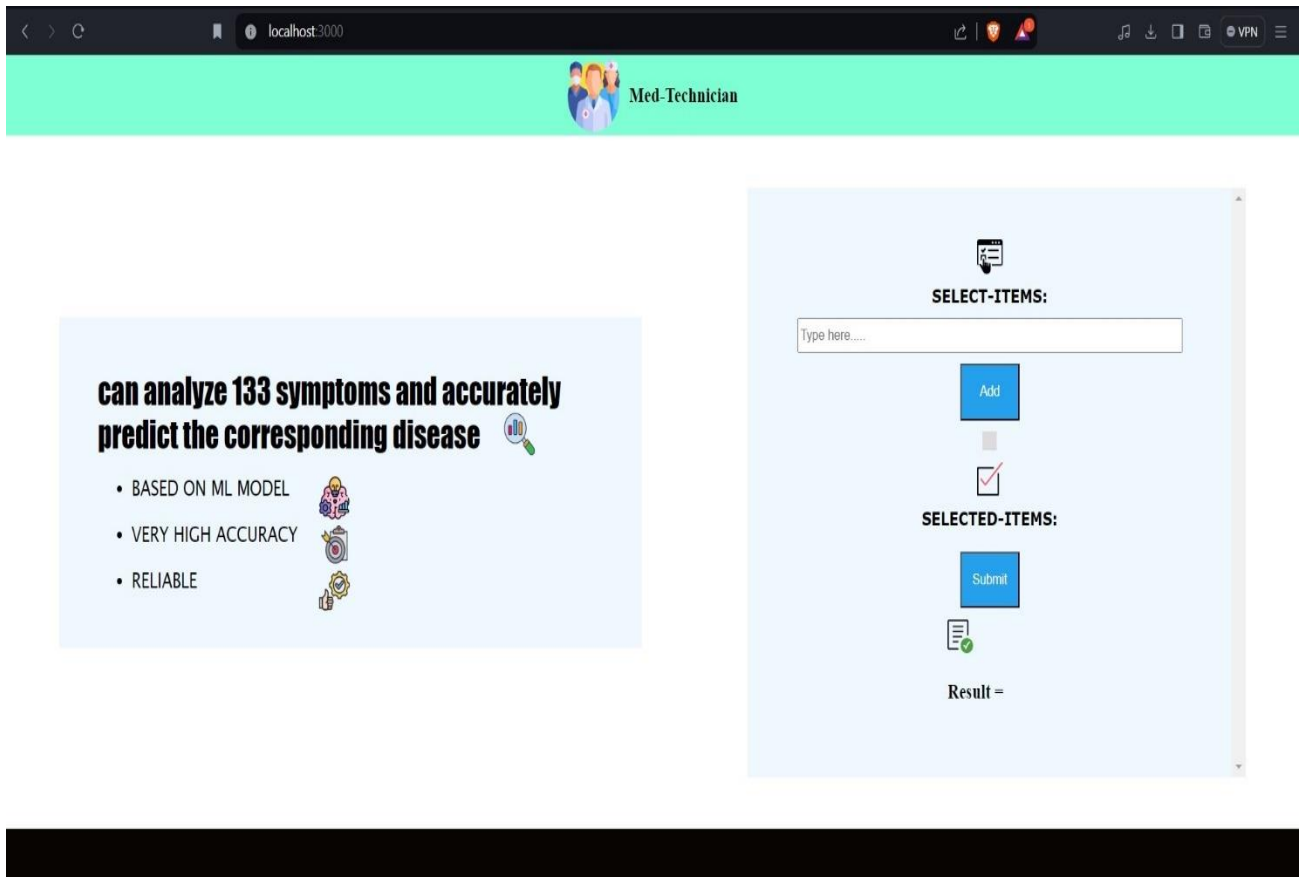# CHAPTER – 7

SCREENSHOTS

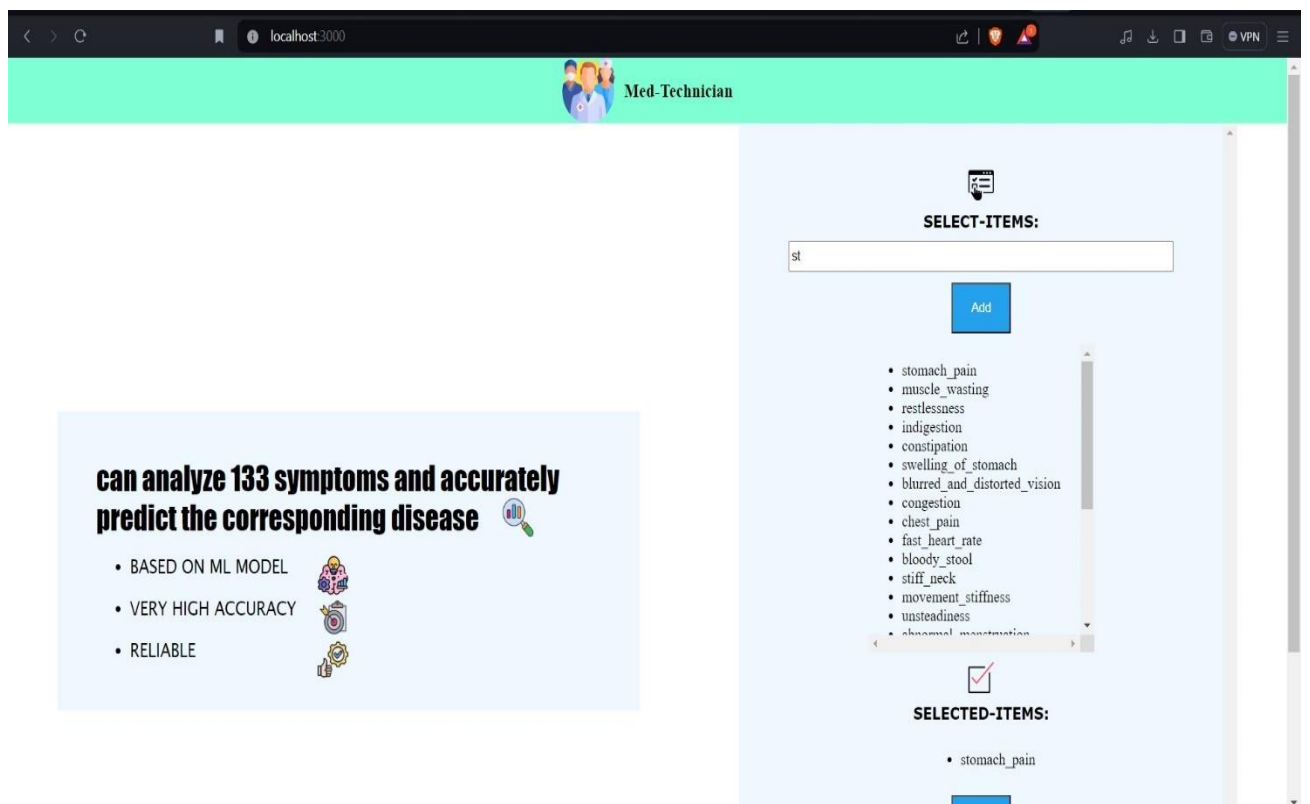Figure 7.1: Frontend of website
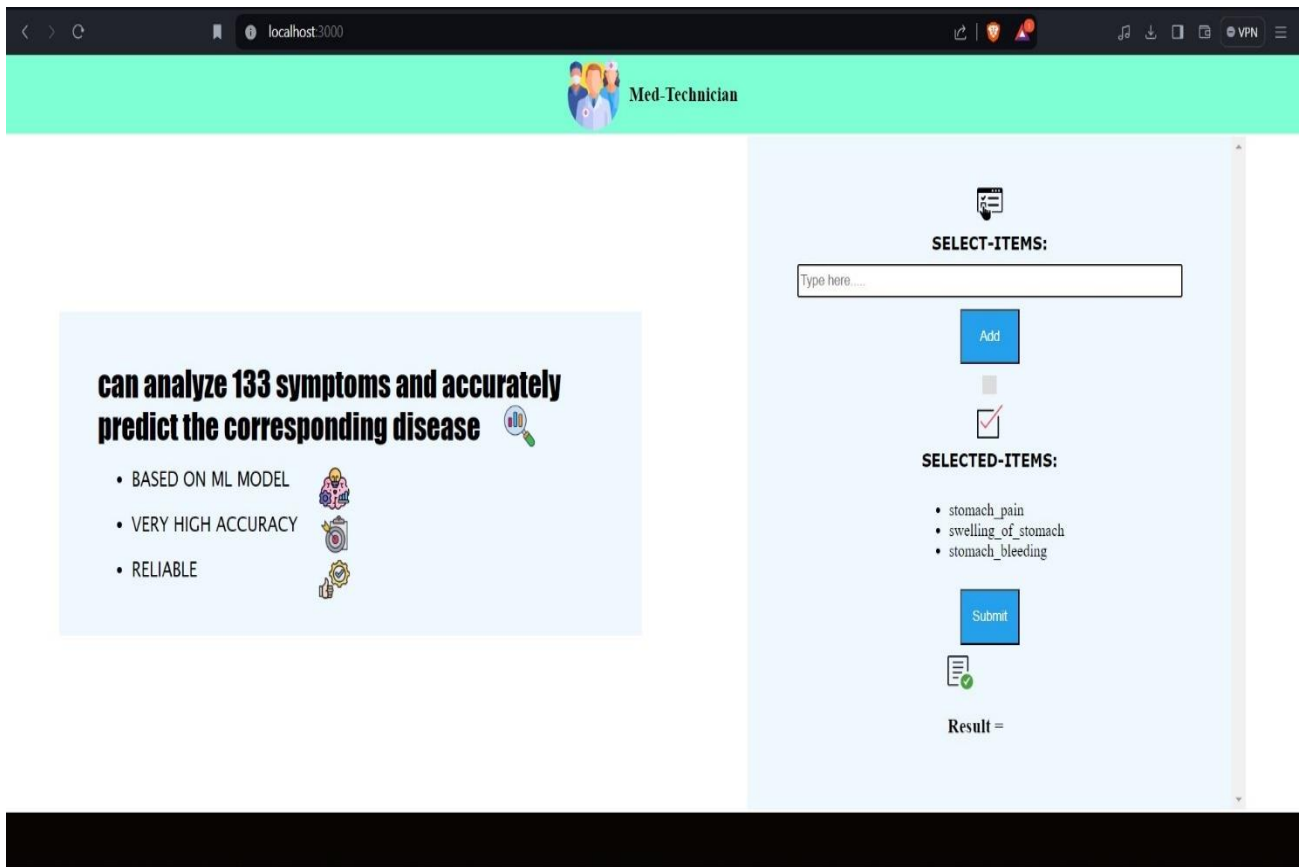
Figure 7.1: Selecting the symptoms

Figure 7.2: Submitting the symptoms
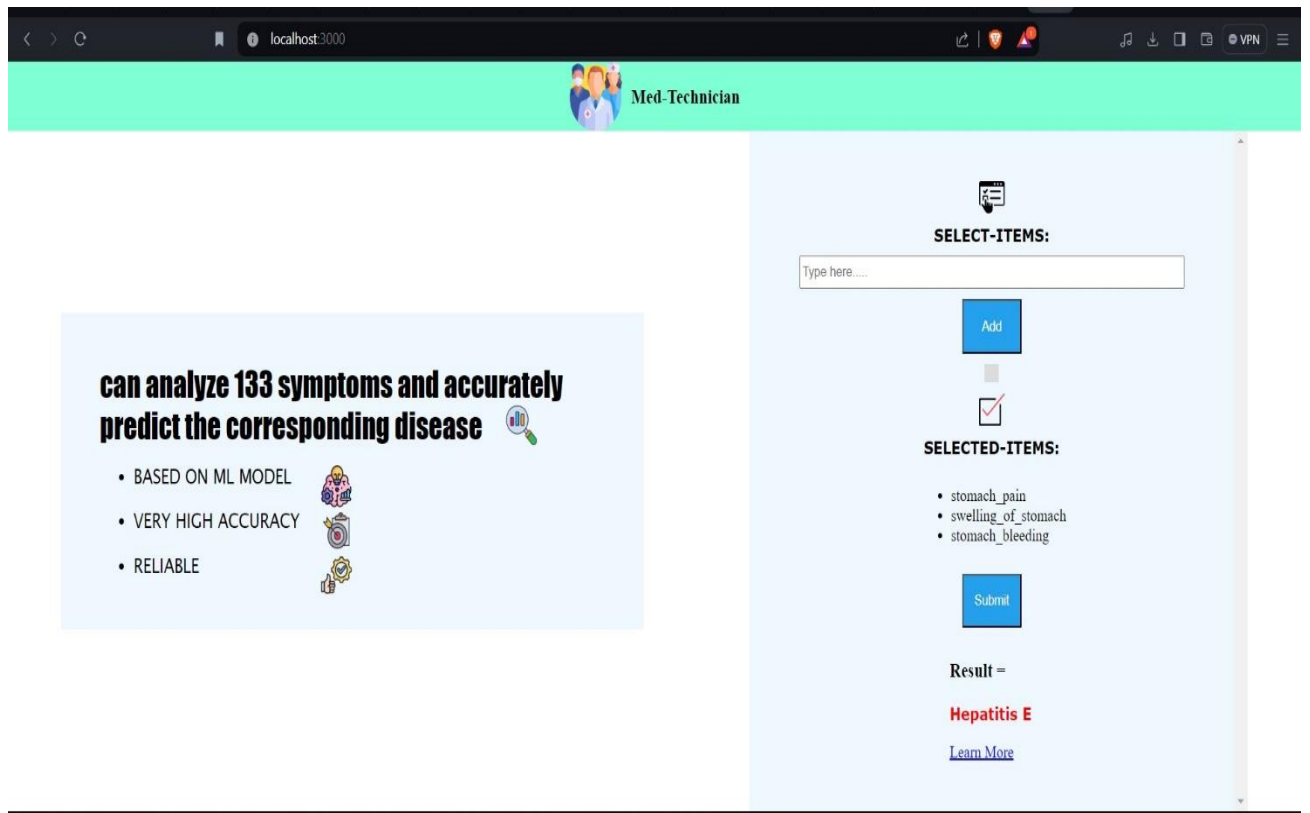
Figure 7.2: Getting the Predicted Disease and clicking on learn more

Figure 7.3: Redirecting to the Wikipedia page of that disease

CHAPTER - 8

CONCLUSION AND FUTURE SCOPE

The future of disease prognosis using the Random Forest classifier holds great promise due to several factors. Firstly, advancements in data collection techniques, such as wearable devices, electronic health records, and genetic testing, are generating vast amounts of diverse and high-dimensional healthcare data. Random Forest, with its ability to handle large datasets with high dimensionality and mixed data types, is well-suited for analyzing such complex healthcare data.

Secondly, the integration of Random Forest with other machine learning techniques, such as deep learning and ensemble methods, can further enhance its predictive accuracy and robustness. By combining the strengths of different algorithms, more comprehensive and accurate disease prognosis models can be developed.

Additionally, ongoing research in feature selection, model interpretability, and explainable AI techniques will improve the transparency and interpretability of Random Forest models, making them more clinically relevant and trustworthy for healthcare professionals.

Furthermore, the application of Random Forest in personalized medicine and precision healthcare, where treatment decisions are tailored to individual patient characteristics, holds significant potential. By leveraging patient-specific data and clinical variables, Random Forest models can assist in early detection, risk stratification, and personalized treatment planning for various diseases.

Overall, the future of disease prognosis using Random Forest classifier is bright, with opportunities for further innovation and advancements in healthcare analytics and personalized medicine.

BIBLIOGRAPHY

[1] Chen, C. (2018). Predicting heart disease using machine learning algorithms. International Journal of Engineering Research and Technology, 11(8), 1219-1223..

[2] Ding, Y., Tang, J., & Guo, F. (2015). A random forest model for predicting diabetes mellitus. International Journal of Bioinformatics Research and Applications, 11(5), 472-485.Recognition," ICLR, 2015.

[3] Ghasemi, A., Zahediasl, S., & Azimzadeh, P. (2012). A novel hybrid naive Bayes, K-nearest neighbors and random forests classifier for the heart disease diagnosis. Journal of Medical Systems, 36(5), 3403-3410. SISAP '10, 2010, p. 101.

[4] Sharma, V., Kumar, R., & Sharma, P. (2019). Predictive modeling for breast cancer using random forest classifier. International Journal of Computer Sciences and Engineering, 7(6), 596-600

[5] Wang, Y., Liu, J., & Jin, Z. (2018). Application of random forest algorithm in early warning of infectious diseases. Journal of Computer Applications, 38(1), 12-16.[6] Y. Chen, Z. Lin, and X. Zhao.

[6] Zheng, H., Jin, B., & Li, Z. (2017). A prediction model for infectious diseases based on random forest and multilayer perceptron. Journal of Medical Imaging and Health Informatics, 7(4), 902-906

[7] Zhou, X., Wu, L., & Zhang, D. (2016). A hybrid method for predicting heart disease using random forest and decision tree. International Journal of Hybrid Information Technology, 9(12), 65-76.

[8] Zhu, J., Xue, Z., & Luo, Y. (2019). A novel method for disease prediction based on random forest. Frontiers in Genetics, 10, 268.

[9] G. Mercier and M. Lennon, "Support vector machines for hyperspectral image classification with spectral-based kernels," in IGARSS 2003. 2003 IEEE

APPENDIX A: TOOLS AND TECHNOLOGIES

- HTML (Hypertext Markup Language): Markup language for creating web pages and web applications.

- CSS (Cascading Style Sheets): Style sheet language for describing the presentation of HTML and XML documents.

- JavaScript (JS): High-level, interpreted scripting language used for adding interactivity to web pages.

- Node.js: Server-side JavaScript runtime environment for building scalable network applications.

- Express.js: Minimal and flexible Node.js web application framework for building web applications and APIs.

- Python: High-level, general-purpose programming language known for its readability and versatility.

- Jupyter Notebook: Open-source web application for creating and sharing documents containing live code, equations, visualizations, and narrative text.

- Random Forest: Ensemble learning method for classification, regression, and other tasks using multiple decision trees.