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

G. CHANDU 2453217480252

N. VISHAL 245321748064

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



**NEIL GOGTE INSTITUTE OF TECHNOLOGY**

A Unit of Keshav Memorial Technical Education (KMTES)

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

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), G. CHANDU(245321748025), N. VISHAL(245321748064) 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

Mr. P. Nageswara Rao

Assistant Professor

Head of Department

Dr. T. Prem Chander

Associate Professor

External

 NEIL GOGTE INSTITUTE OF TECHNOLOGY

A Unit of Keshav Memorial Technical Education (KMTES)

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

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

G. CHANDU 245321748025

N. VISHAL 245321748064

ACKNOWLEDGEMENT

I am happy to express our deep sense of gratitude to the principal of the college Dr. R. Shyam Sunder, Professor, Neil Gogte Institute of Technology, for having provided us with adequate facilities to pursue our project.

I would like to thank, Dr. T. Prem Chander, Head of the Department, CSE(AIML), Neil Gogte

Institute of Technology, for having provided the freedom to use all the facilities available in the

department, especially the laboratories and the library.

I would also like to thank my internal guide Mr. P. Nageswara Rao, Assistant Professor for his technical guidance & constant encouragement.

I sincerely thank my seniors and all the teaching and non-teaching staff of the Department of Computer Science & Engineering for their timely suggestions, healthy criticism and motivation during this work.

Finally, I express my immense gratitude with pleasure to the other individuals who have either directly or indirectly contributed to our need at the right time for the development and success of this work.

I

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.

II

TABLE OF CONTENTS

CHAPTER TITLE PAGE NO.

ACKNOWLEDGEMENT I

ABSTRACT II

LIST OF FIGURES V

LIST OF TABLES V

1. INTRODUCTION

1.1. PROBLEM STATEMENT 1

1.2. MOTIVATION 1

1.3. SCOPE 1

1.4. OUTLINE 2

2. LITERATURE SURVEY

2.1. EXISTING SYSTEM 3

2.2. PROPOSED SYSTEM 3

3. SOFTWARE REQUIREMENT SPECIFICATION

3.1. OVERALL DESCRIPTION 4

3.2. OPERATING ENVIRONMENT 4

3.3. FUNCTIONAL REQUIREMENTS 4 - 5

3.4. NON – FUNCTIONAL REQUIREMENTS 5 – 7

II

4. SYSTEM DESIGN

4.1. USE-CASE DIAGRAM 8 – 9

4.2. CLASS DIAGRAM 10

4.3. SEQUENCE DIAGRAM 11

5. IMPLEMENTATION

5.1. SAMPLE CODE 12 – 21

6. TESTING

6.1. TEST CASES 22 – 23

7. SCREENSHOTS 24 - 38

8. CONCLUSION AND FUTURE SCOPE 29

BIBLIOGRAPHY 30

APPENDIX A: TOOLS AND TECHNOLOGY 31

IV

List of Figures Figure No.

1.

2.

3.

Name of Figure

Use case Diagram

Class Diagram

Sequence Diagram

Page No.

8 - 9

10

11

List of Tables

Table No.

1.

Name of Table

Testcases

Page No.

22 - 23

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 : Front End : Back End : Database :

Hardware Requirements: Processor : Speed : RAM :

Hard Disk :

Windows 7 (Min)

Html, CSS, JS

Nodejs ,Express Microsoft Excel

Intel Pentium® Dual Core Processor (Min) 2.9 GHz (Min)

2 GB (Min)

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:

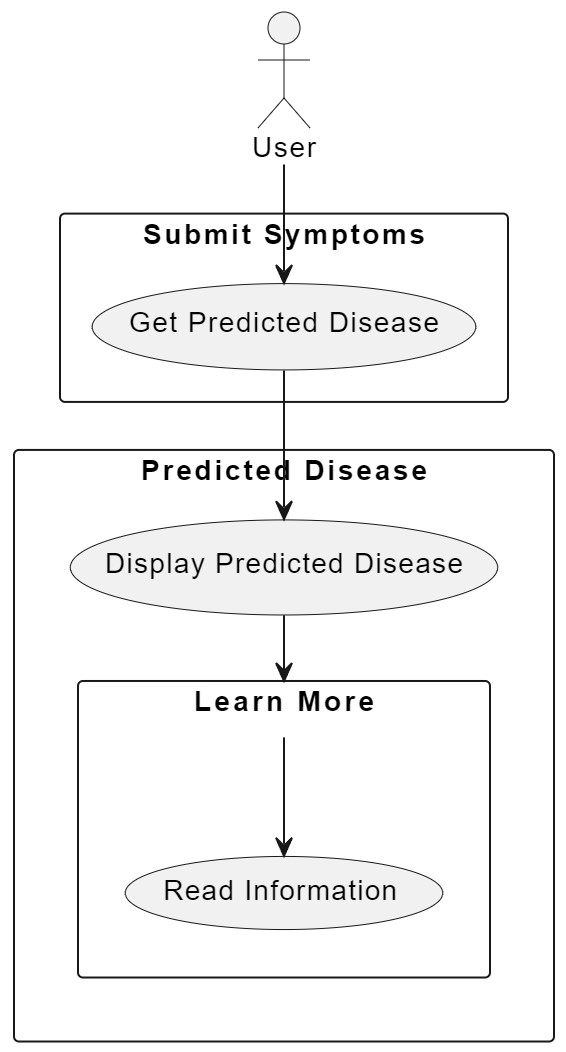


Fig 4.1: Use case diagram for User

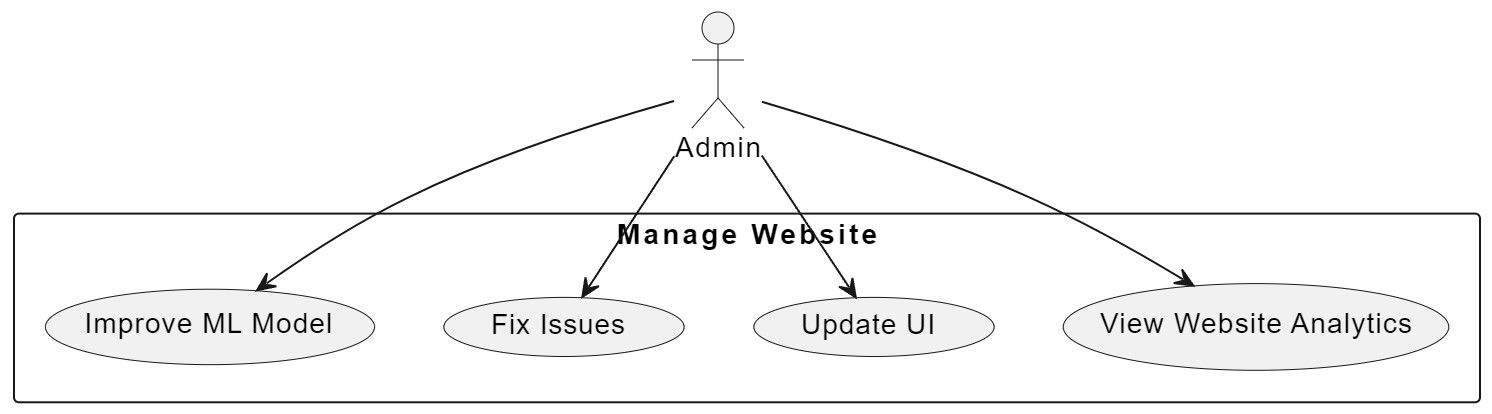


Fig 4.2: Use case diagram for Admin

Class Diagram:

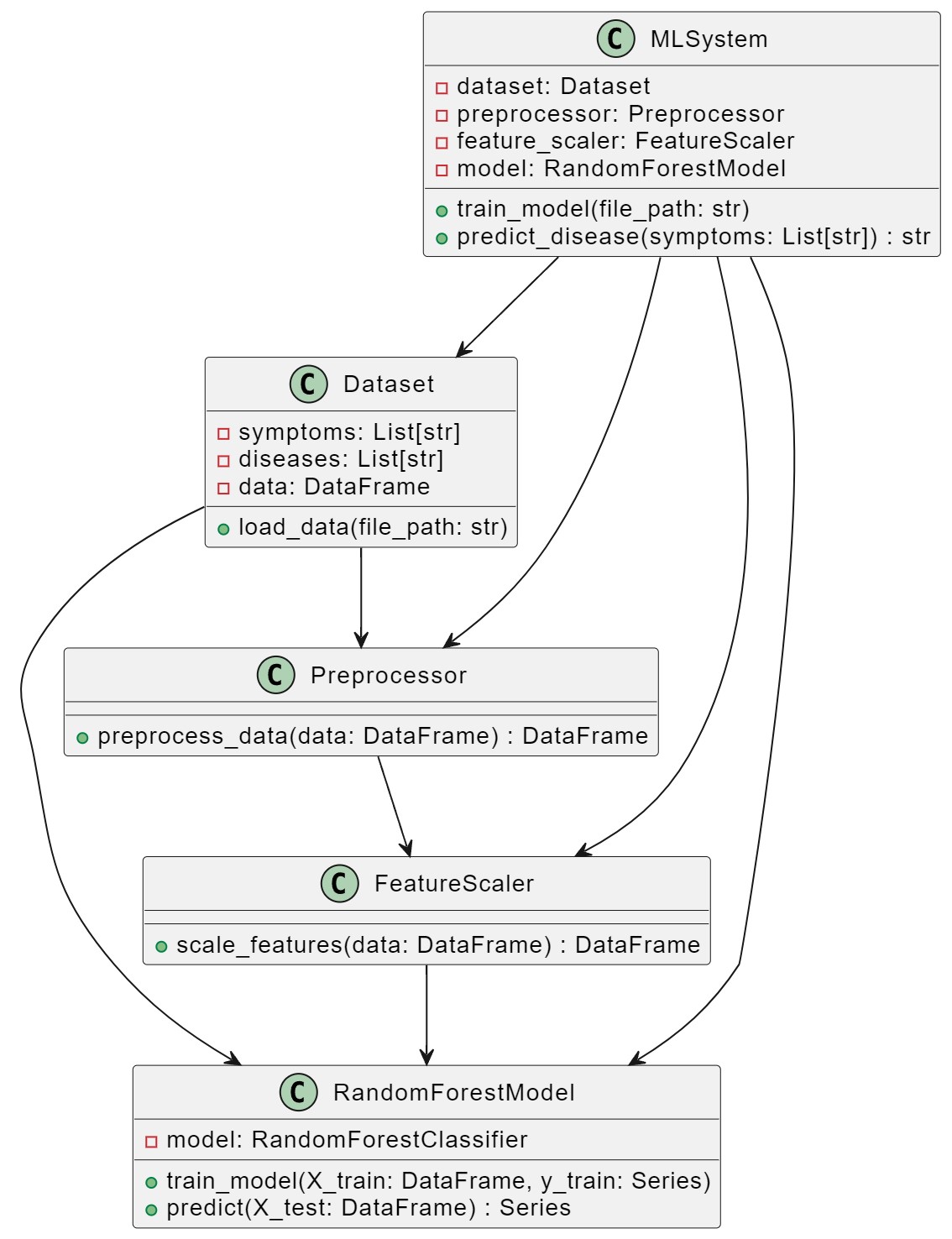


Fig.4.3 : Class diagram for Model Training.

Sequence Diagram:

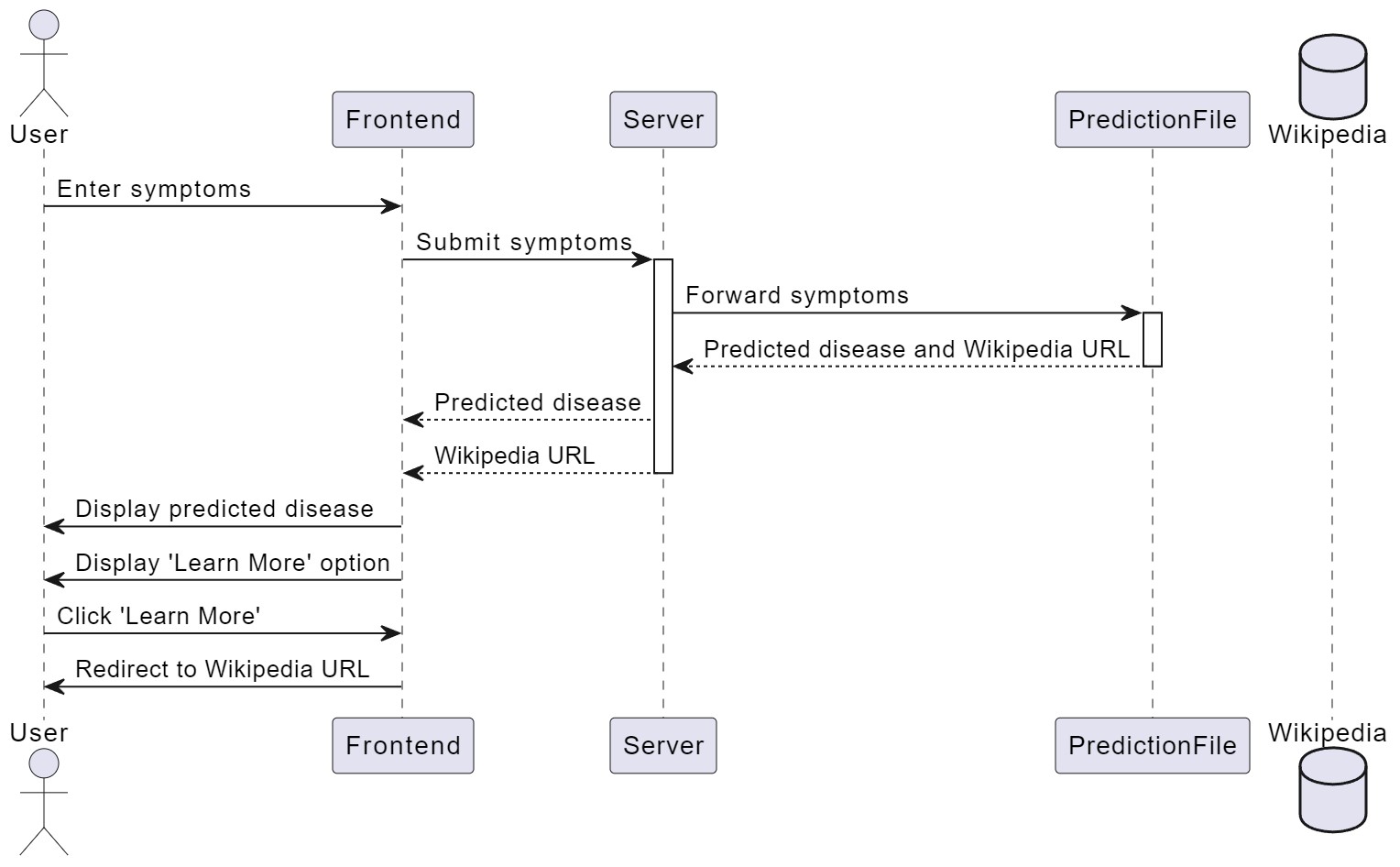


Fig4.3 : Sequence Diagam for Disease Prediction.

CHAPTER – 5

IMPLEMENTATION

5.1 SAMPLE CODE

**Model Training Code :-**

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

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

from sklearn.model\_selection import train\_test\_split

X = df.iloc[:,:132]

y = df['prognosis']

from sklearn.preprocessing import StandardScaler

std = StandardScaler()

X = std.fit\_transform(X)

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

from sklearn.ensemble import RandomForestClassifier

RFC = RandomForestClassifier()

RFC.fit(X\_train,y\_train)

RandomForestClassifier()

y\_pred = RFC.predict(X\_test)

y\_pred

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

from sklearn.metrics import accuracy\_score

print(accuracy\_score(y\_test,y\_pred)\*100)

100.0

y\_train\_pred = RFC.predict(X\_train)

print(accuracy\_score(y\_train,y\_train\_pred)\*100)

100.0

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%

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:

Test Case Name:

Purpose:

Input:

Expected Result:

Actual Result:

Failure

1

Required Software Testing

To check whether the required Software is installed on the systems

Enter python command

Should Display the version number for the python

Displays python version

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:

Test Case Name:

Purpose:

Input:

Expected Result:

Actual Result:

Failure

2

Programs Integration Testing

To ensure that all the modules work together

All the modules should be accessed.

All the modules should be functioning properly.

All the modules should be functioning properly.

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:

Test Case Name:

Purpose:

Input:

Expected Result:

Actual Result:

Failure

3

Collect Dataset and Load the Dataset

Check Dataset is collected, and the data is stored

Provide Dataset as input

Dataset is collected and view the Dataset and store the Dataset

Load the Dataset and view the Dataset and store

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:

Test Case Name:

Purpose:

Input:

Expected Result:

Actual Result:

Failure

4

Disease Prediction

Disease Prediction using RandomForestClassifier

Provide dataset and input the symptoms

After Evaluation I get the Disease Predicted

After Evaluation I get the Disease Predicted

If the data is not Evaluated, it does not show anything

Table 6.1.4 Disease Prediction

CHAPTER – 7

SCREENSHOTS

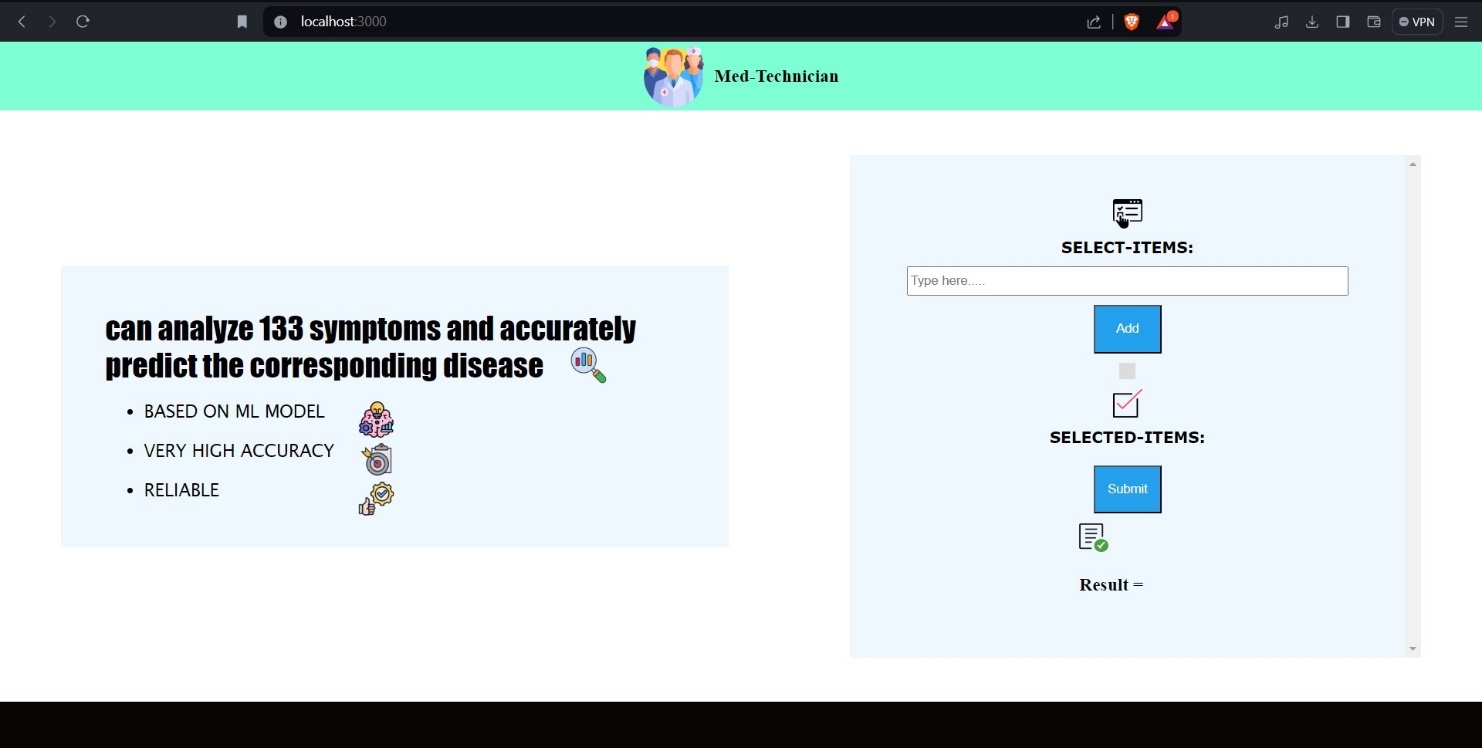


Figure 7.1: Frontend of website

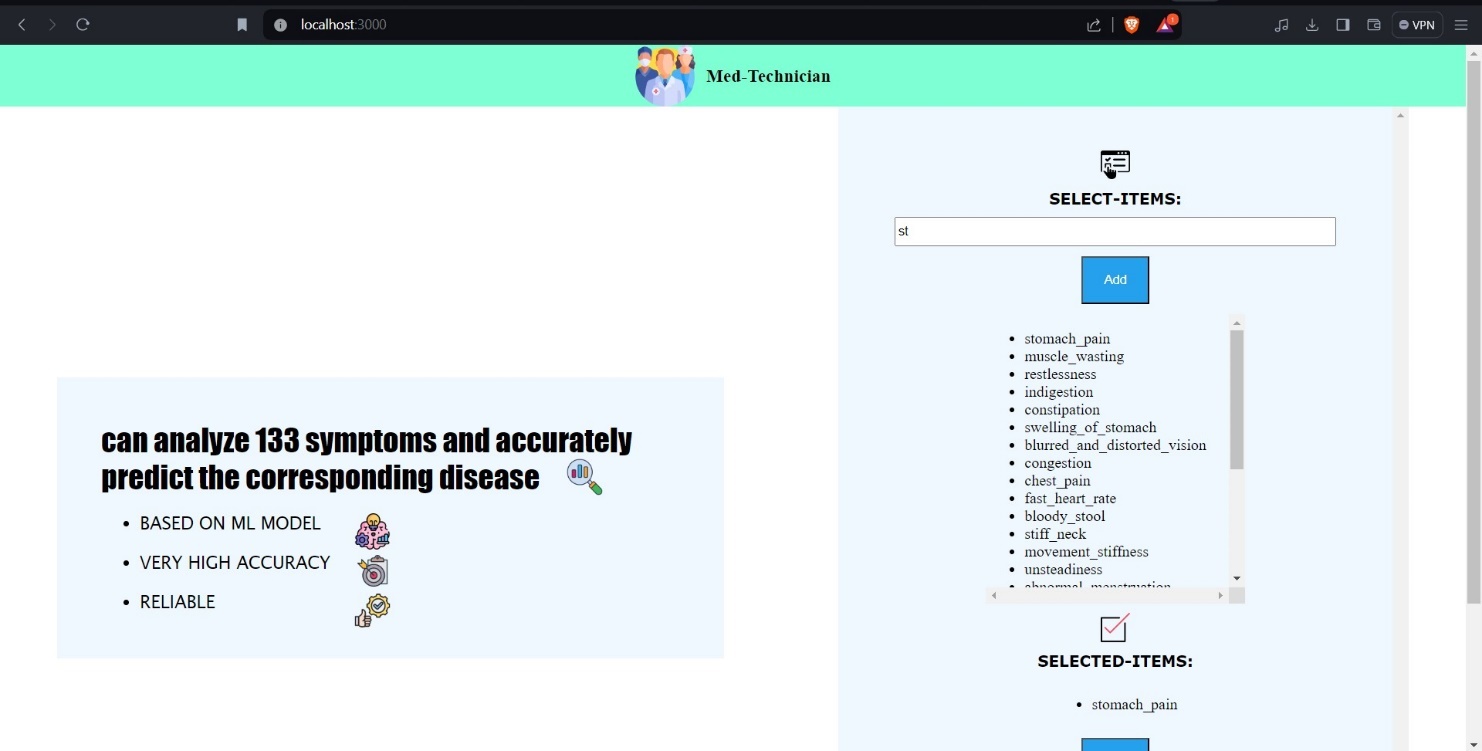


Figure 7.1: Selecting the symptoms

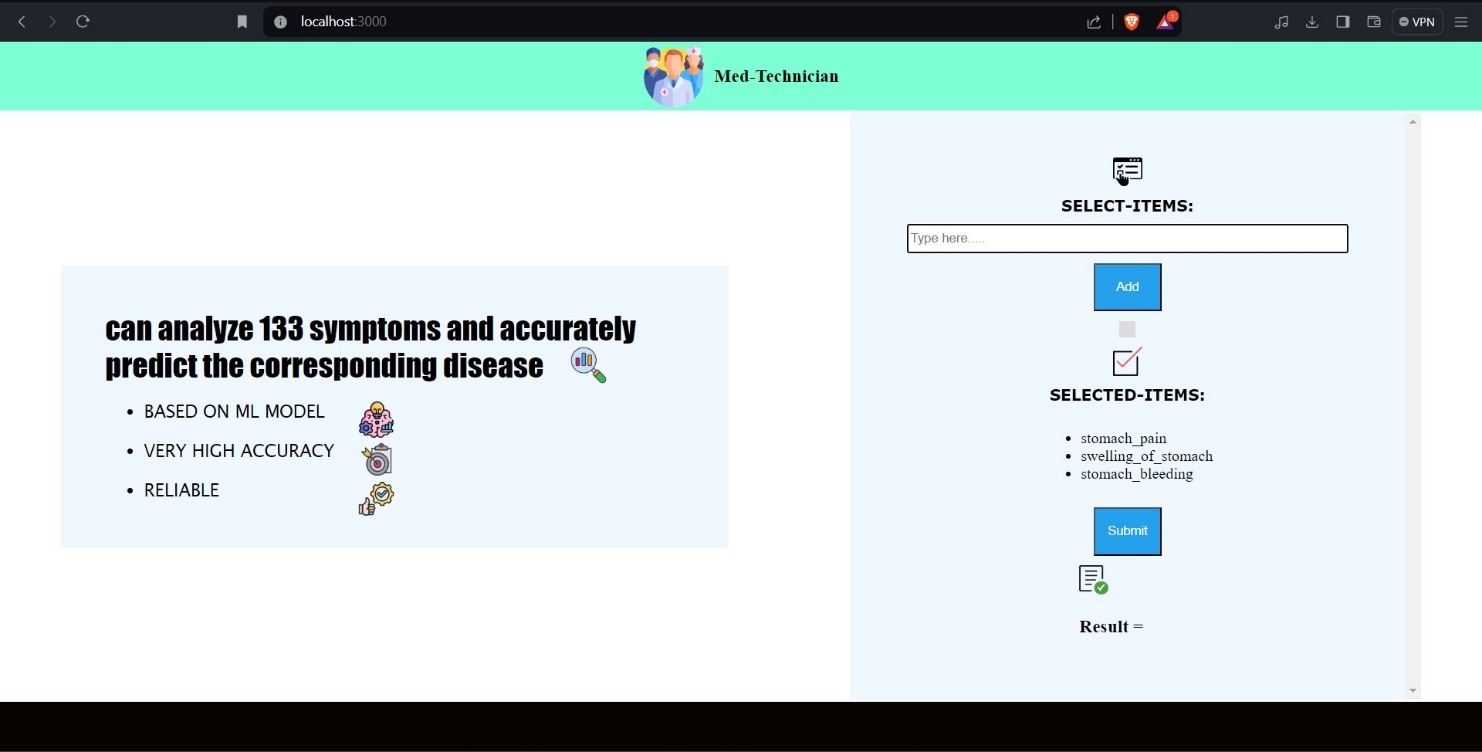


Figure 7.2: Submitting the symptoms

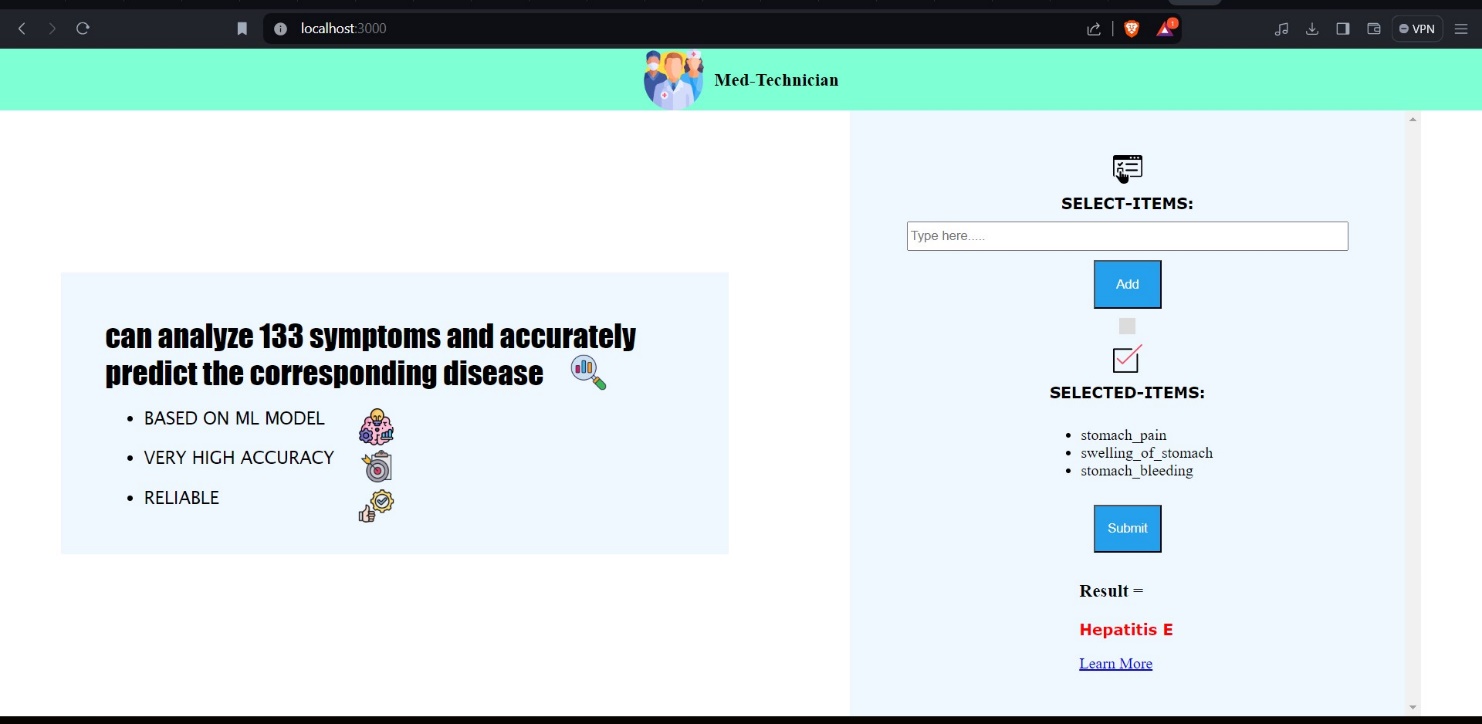
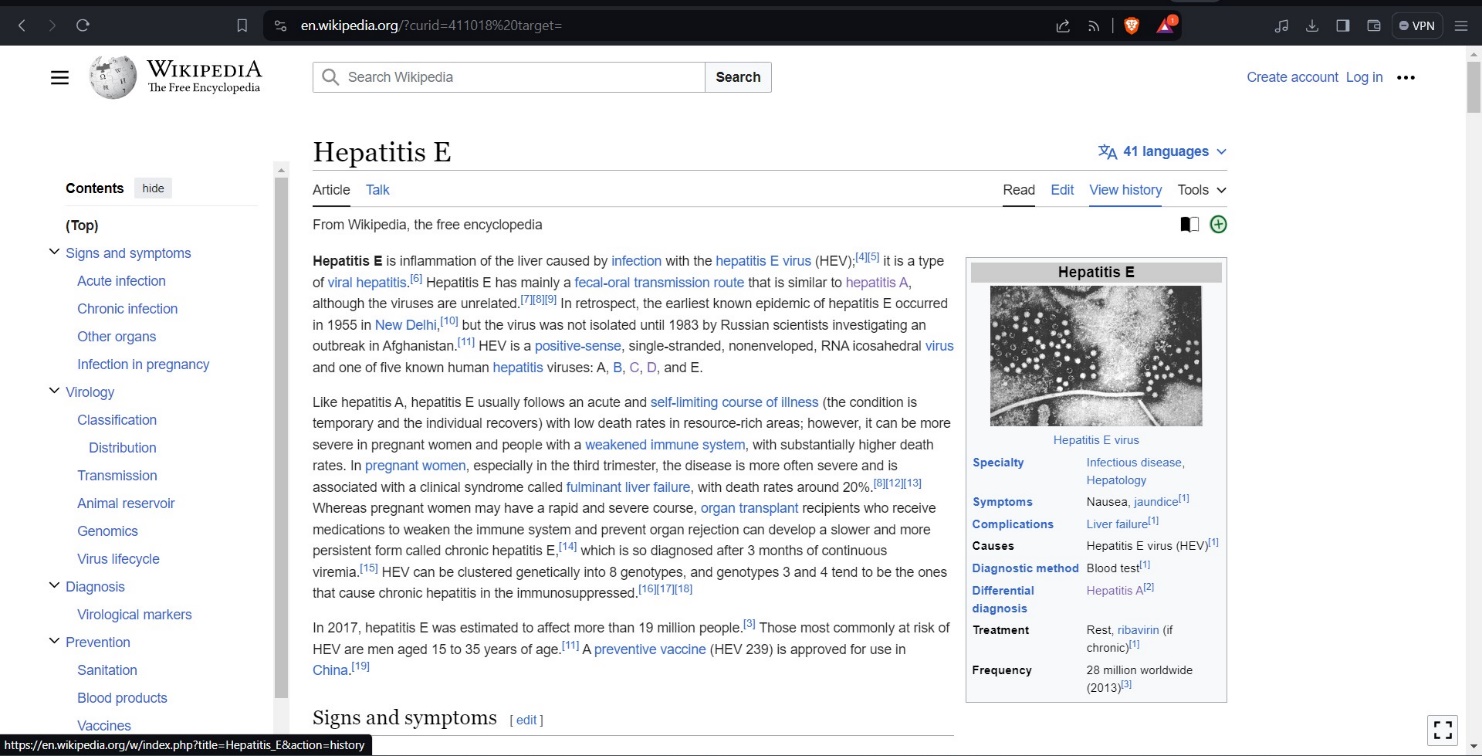


Figure 7.2: Getting the Predicted Disease and clicking on learn more



**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.