

A Mini Project Report

on

**SELF DIAGNOSABLE HUMAN DISEASE PREDICTION
USING MACHINE LEARNING**

Submitted in partial fulfillment of the
Requirements for the award of degree of

Bachelor of Technology

in

Computer Science and Engineering

by

**GAJAM NIKHIL
(19H61A05D5)**

**EEDUNOORI KRUTHIK REDDY
(19H61A05D1)**

**GADIPE DOLLY
(19H61A05D4)**

Under the Guidance of

**Mr. G. Balram
(Asst. Professor)
Department of CSE**



**Department of Computer Science and Engineering
ANURAG GROUP OF INSTITUTIONS
(Formerly CVSR College of Engineering)**

**(An Autonomous Institution, Approved by AICTE and NBA Accredited)
Venkatapur (V), Ghatkesar (M), Medchal(D)., T.S-500088
(2019-2023)**



Anurag Group of Institutions

An Autonomous Institution
(Formerly CVSR College of Engineering)
Venkatapur (V), Ghatkesar (M), Ranga Reddy (Dist.)
Ph: 08499953666, 08499963666. www.anurag.edu.in

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that the project entitled “**SELF DIAGNOSABLE HUMAN DISEASE PREDICTION USING MACHINE LEARNING**” being submitted by **GAJAM NIKHIL, EEDUNOORI KRUTHIK REDDY, GADIPE DOLLY** bearing the Hall Ticket number **19H61A05D5, 19H61A05D1, 19H61A05D4** in partial fulfillment of the requirements for the award of the degree of the **Bachelor of Technology in Computer Science and Engineering** to **Anurag Group of Institutions (Formerly CVSR College of Engineering)** is a record of bonafide work carried out by them under my guidance and supervision from June 2022 to October 2022.

The results presented in this project have been verified and found to be satisfactory. The results embodied in this project report have not been submitted to any other University for the award of any other degree or diploma.

Internal Guide
Mr. G. Balram
(Asst. Professor)
Department of CSE

External Examiner

Dr.G.Vishnu Murthy,
Professor & Head, Dept. of CSE

ACKNOWLEDGEMENT

It is our privilege and pleasure to express profound sense of respect, gratitude and indebtedness to our guide **Mr. G. Balram**, Asst Professor, Dept. of Computer Science and Engineering, Anurag Group of Institutions (Formerly CVSR College of Engineering), for his indefatigable inspiration, guidance, cogent discussion, constructive criticisms and encouragement throughout this dissertation work.

We express our sincere gratitude to **Dr. G. Vishnu Murthy, Professor & Head**, Department of Computer Science and Engineering, Anurag Group of Institutions (Formerly CVSR College of Engineering), for his suggestions, motivations, and co-operation for the successful completion of the work.

We extend our sincere thanks to **Dr. V. Vijaya Kumar, Dean**, Research and Development, Anurag Group of Institutions, for his encouragement and constant help.

GAJAM NIKHIL
(19H61A05D5)

EEDUNOORI KRUTHIK REDDY
(19H61A05D1)

GADIPE DOLLY
(19H61A05D4)

DECLARATION

We hereby declare that the project work entitled “**SELF DIAGNOSABLE HUMAN DISEASE PREDICTION USING MACHINE LEARNING**” submitted to the **Anurag Group of Institutions(Formerly CVSR College of Engineering)** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology (B. Tech)** in Computer Science and Engineering is a record of an original work done by us under the guidance of **Mr. G. Balram, Assistant Professor** and this project work have not been submitted to any other university for the award of any other degree or diploma.

GAJAM NIKHIL
(19H61A05D5)

EEDUNOORI KRUTHIK REDDY
(19H61A05D1)

GADIPE DOLLY
(19H61A05D4)

Date:

ABSTRACT

Self Diagnosable Human Disease Prediction using Machine Learning (ML) is a system which predicts the disease based on the information provided by the user. It also predicts the disease of the patient or the user based on the information or the symptoms he/she enter into the system and provides the accurate results based on that information. If the patient is not much serious and the user just wants to know the type of disease, he/she has been through. It is a system which provides a way to find out the disease using this prediction. Now a day's health industry plays major role in curing the diseases of the patients so this is also some kind of help for the health industry to tell the user and also it is useful for the user in case he/she doesn't want to go to the hospital or any other clinics, so just by entering the symptoms the user can get to know the disease he/she is suffering from and the health industry can also get benefit from this system by just asking the symptoms from the user and entering in the system and in just few seconds they can tell the exact and up to some extent the accurate diseases. This Self Diagnosable Human Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with Tkinter Interface for it and also using the dataset which is taken from a study conducted at Columbia University which is readily available, using that we will predict the disease. Today's doctors use a variety of scientific technology and methodologies for both the detection and diagnosis of numerous fatal diseases as well as ordinary illnesses. Correct and thorough diagnosis is always responsible for a successful course of treatment. Doctors may sometimes be able to make accurate decisions while diagnosing the disease of a patient, therefore disease prediction systems which use machine learning algorithms assist in such cases to get accurate results. Because of the competitive environment of economic development and the prevalence of general disease, the project disease prediction using machine learning was created to combat general disease in its earlier stages involved so much that he/she is not concerned about health according to research there are 40% peoples how ignores about general disease which leads to harmful disease later. The main reason of ignorance is laziness to consult a doctor and time concern the peoples have involved themselves so much that they have no time to take an appointment and go to the doctor which later results into fatal disease. According to research there are 70% peoples in India suffers from general disease and 25% of peoples face death due to early ignorance of the symptoms.

CONTENTS

S.NO	PAGE NO
1. Introduction	1
1.1. Motivation	2
1.2. Problem Definition	2
1.3. Objective of the Project	2
2. Literature Survey	3
3. Analysis	5
3.1. Existing System	5
3.2. Proposed System	5
3.3. Software Requirement Specification	6
3.3.1 Purpose	6
3.3.2 Scope	6
3.3.3 Overall Description	6
4. Design	8
4.1. UML Diagrams	8
4.1.1. Use Case	8
4.1.2. Class Diagram	9
4.1.3 Activity Diagram	10
4.1.4 Sequence Diagram	11
5. Implementation	12
5.1. Modules	12
5.2. Module Description	12

5.3 Introduction to Technologies Used	13
5.4 Sample Code	15
6. Test Cases	21
7. Screenshots	22
8. Conclusion	24
9. Future Enhancement	25
10. Bibliography	26

1. INTRODUCTION

Now-a-days, many people having various diseases due to environmental conditions and their living habits. So the prediction of disease at earlier stage becomes an important task. The normal medical check up is a process where the patient has to visit a doctor and undergo many medical tests, and then come to a conclusion. This process is very time consuming. To save time required for the initial process of diagnosis symptoms, this project proposes an automated disease prediction system that relies on user input. This Self Diagnosable Human Disease Prediction Using Machine Learning is completely done with the help of Machine Learning and Python Programming language with Tkinter Interface for it and also using the dataset which is taken from a study conducted at Columbia University which is readily available, using that we will predict the disease. Today's doctors use a variety of scientific technology and methodologies for both the detection and diagnosis of numerous fatal diseases as well as ordinary illnesses. Correct and thorough diagnosis is always responsible for a successful course of treatment. Doctors may sometimes be able to make accurate decisions while diagnosing the disease of a patient, therefore disease prediction systems which use machine learning algorithms assist in such cases to get accurate results. Because of the competitive environment of economic development and the prevalence of general disease, the project disease prediction using machine learning was created to combat general disease in its earlier stages involved so much that he/she is not concerned about health according to research there are 40% peoples how ignores about general disease which leads to harmful disease later. The main reason of ignorance is laziness to consult a doctor and time concern the peoples have involved themselves so much that they have no time to take an appointment and go to the doctor which later results into fatal disease. According to research there are 70% peoples in India suffers from general disease and 25% of peoples face death due to early ignorance of the symptoms.

1.1 MOTIVATION

People are looking online for health information regarding diseases, diagnoses, and different treatments. If a recommendation system is made for users then it will save a lot of time. The main motive to develop this project is that a user can sit at their convenient place and have a check up of their health the UI is designed in a simple way that everyone can easily operate on it and can have a check up.

1.2 PROBLEM DEFINITION

Now-a-days in Health Industry there are various problems related to machines or devices which will give wrong or unaccepted results. We are developing a project that will make accurate predictions based on the data that is already accessible to the machine as well as the information provided by the user in order to prevent those outcomes and obtain the correct and desired results. The health sector is a very large industry with a lot of unfinished things and is still underdeveloped in terms of knowledge and information. So, with the help of all those algorithms, techniques and methodologies we have done this project which will help the people who are in the need of a health check up.

So the problem here is that many people will go to hospitals or clinics to know how is their health and how much they got improved in the given days, but they have to travel to get to know these answers and sometimes the patients may or may not get the results based on various factors such as doctor might be on leave or some weather problem. So, he might not have come to the hospital and many more reasons will also be there. So, to avoid all these reasons and confusion we are making a project which will help all those person's and all the patients who are in need to know the condition of their health, and at sometimes if the person has been observing few symptoms and he/she is not sure about the disease he/she has encountered with. So, this will lead to various diseases in future. In order to avoid that and get to know the disease in early stages of the symptoms this disease prediction system will help a lot of people ranging from children to teenagers to adults and also the senior citizens.

1.3 OBJECTIVE OF THE PROJECT

There is a need to study and make a system which will make it easy for an end user to predict the diseases without visiting physician or doctor for diagnosis. The Predictions Accuracy will Increase using Machine Learning Naive Bayes Algorithm.

2. LITERATURE SURVEY

In the paper called “Disease Prediction System using data mining techniques” the author has discussed about the data mining techniques like association rule mining, classification, clustering to analyse the different kinds of heart disease based problems. The database used contains collection of records, each with a single class label. A classifier performs a simple and clear definition for each class that can be used to classify successive records. The data classification depends on the algorithm called MAFLA which causes accuracy, the info is calculable exploitation entropy primarily based cross validations and partition techniques and also the results are compared with other outputs. C4.5 algorithmic rule is employed because the coaching algorithmic rule to indicate rank of attack with the choice tree. The heart disease problem information is clustered mistreatment which is the K-means clump algorithmic rule, which will remove the data applicable to heart attack from the database. Some limitations are measured by the measurements like, time complexity is more due to DFS traversal, C4.5- Time complexity increases while searching for insignificant branches and lastly no precautions are defined.

In the paper "A study on data mining prediction techniques in healthcare sector" 21 fields that mentioned are, information Discovery method (KDD) is that the method of adjusting the low level data into high-level knowledge. The repetition method consists of the subsequent steps information cleansing, information integration, information choice, information transformation, data processing, Pattern analysis, Knowledge is the process.

Healthcare data processing prediction will support data processing techniques which are as follows: Neural network, Bayesian Classifiers, call tree, Support Vector Machine. The paper states that comparative study of various aid predictions, Study of information mining techniques and tools for prediction of cardiovascular disease, numerous cancers, and diabetes, disease and medicine conditions. Some limitations are if attributes are not related then Decision tree prediction is less accurate and ANN is computationally intensive to train also it will not lead to specific conclusion. So, we can use Bayes algorithm.

The paper “Predicting Disease by Using Data Mining Based on Healthcare information System” published in the year 2016 applies the information mining process to predict high blood pressure from patient medical records with eight alternative diseases. The data was taken from a true world health care system information containing medical records. Under sampling technique has been applied to come up with coaching knowledge sets, and data processing tool will not generate the Naive Bayesian and J-48 classifiers created to improve the prediction performance, and rough set tools were wanted to scale back the ensemble supported the concept of second order approximation.

Experimental outputs showed a bit improvement of the ensemble approach over pure Naive Bayesian and J-48 in accuracy, sensitivity and F-measure. Initially they used classification called ensemble the classifier and so the reduction of Ensemble Classifiers is employed. But the choice trees generated by J-48 is typically lacking within the levelling therefore the overall improvement of victimization ensemble approach is a small amount.

The paper "An approach to devise an Interactive software solution for smart health prediction using data mining" published in the year 2018 aims in developing a computerized system to check and maintain your health by knowing the symptoms. It has a symptom checking module which actually defines our body structure and gives us liability to select the affected area and checkout the symptoms. Technologies which are implemented in this paper are: The front end is developed with using HTML, Java Script and CSS. The back end is developed using MySQL which is used to design the database queries. This paper also contains the information of testing like Alpha testing which is performed at server side or we can say at the developer's end, this is an actual testing done with potential users or as an independent testing process at server end. And the Beta testing is done after performing alpha testing, versions of a system or software known as beta versions are given to a specific audience outside the programming team to check. Only the limitation of this paper is that it suggests that only the award winning doctors and not the nearby doctors to the patient who is using this algorithm.

3. ANALYSIS

3.1 EXISTING SYSTEM

Prediction with the traditional methods and models involves various risk factors and it consists of various measures of algorithms such as datasets, programs and much more to consider. High risk and Low risk patient classification is made on the basis of the tests that are used in that group. But these models are only used in the clinical sector and not in huge industry sector. So, to say the disease predictions in various health related industries, we have used some of the concepts of machine learning and supervised learning methods to build the predictions system. Existing system can predict the disease but not the sub type of that particular disease. And it cannot predict the condition of the people. Predictions of any disease are not definite and non specific. Existing model gives less accuracy as discussed.

3.2 PROPOSED SYSTEM

The proposed system of disease prediction using machine learning is that we have used many techniques and algorithm and all other various tools to build a system which predicts the disease of the patient using the symptoms and by taking those symptoms we are comparing with the system's dataset that is previously available. By taking those datasets and comparing with the patient's disease we will predict the accurate percentage disease of the patient. The dataset and symptoms go to the prediction model of the system where the data is pre-processed for the future references and then the feature selection is done by the user where he will enter the various symptoms. Then the classification of those data is done with the help of Naive Bayes algorithm. Then the data is given to the recommendation model algorithm, there it shows the risk analysis that is involved in that particular system. And here, we have combined the overall structured and unstructured forms of data for the overall risk analysis that is required for doing the prediction of any disease. Using the structured analysis of data, we can identify the chronic types of disease in a particular region and particular community. In an unstructured analysis of data we select the features automatically with the help of different algorithms and techniques. This system takes symptoms from the user and predicts the disease accordingly based on the symptoms that it takes and also from the previous datasets, it also helps in continuous evaluation of viral diseases, heart rate, blood pressure, sugar level and much more which is in the system and along with other external symptoms its predicts the appropriate and accurate disease.

3.3 SOFTWARE REQUIREMENT SPECIFICATION

3.3.1 PURPOSE

The purpose of making this project called "Human Disease Detection" is to predict the accurate disease of the patient using all their general information's and also the symptoms. Using this information, there we will compare with our previous datasets of the patients and predicts the disease of the patient he/she is been through. If this kind of Prediction is done at the early stages of the disease with the help of this project and all other necessary measures then the disease can be cured and in general this prediction system can also be very useful in health industry to save many lives. If health industry takes this project then the work of the doctors can be reduced and they can easily predict the disease of the patient and save people lives.

The normal purpose of this Disease prediction is to give predictions for the various and generally occurring diseases that when unchecked and sometimes ignored can turns into dangerous disease and cause lot of problems to the patient and as well as their family members. This system will predict the most possible disease based on the symptoms given. The health industry is having information but knowledge is poor and this industry is very vast industry which has lot of work to be done. So, with the help of all those algorithms, techniques and methodologies we have done this project which will help the people who needed check up.

3.3.2 SCOPE

The scope of the project is the integration of medical decision support with computer based patient records could reduce medical diagnosis errors, enhance patient safety and also decrease unwanted practice variation, and improve patient health outcome. This suggestion is promising as data modelling and analysis tools.

3.3.3 OVERALL DESCRIPTION

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and no of times it runs. It is seen as a part of artificial intelligence (AI). Machine learning algorithms build a model based on the sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed by a programmer. Machine learning algorithms are used in many applications, such as email filtering and computer vision, where it is difficult to develop conventional algorithms to perform the needed tasks. In our project we make use of Classification to

create a machine learning model to predict the disease of a User. It uses Naive Bayes Algorithm. To provide a user friendly interface, we implemented it using Tkinter Library. From this page user details and symptoms are collected and pre processed and sent to machine learning model to predict the disease.

SOFTWARE REQUIREMENTS

Operating System : Windows 7, 10 or Higher Versions

Front End : Tkinter

Back End : Python and Modules

Programming Language : Python

HARDWARE REQUIREMENTS

System : Intel Core i3 or above

RAM : 512 MB or above

Hard Disk : 10 GB or above

Input Device : Keyboard and Mouse

Output Device : Monitor or PC

4. DESIGN

4.1 UML DIAGRAMS:

4.1.1 USE CASE DIAGRAM

The Use Case diagram of the project disease prediction using machine learning contains of all the various aspects a general use case diagram requires. This use case diagram shows how to do from starting the model flows from one step to another like he enter into the system then enters all his information like symptoms that goes into the system. compares with the prediction model and if true is predicts the appropriate results otherwise it shows the details where the user if gone wrong while entering the information's and it also shows the appropriate messages for the user to follow. Here the use case diagram of all the entities are linked to each other where the user gets started with the system and in the end output will be presented.

PURPOSE OF USE CASE DIAGRAMS

The purpose of a use case diagram can be described as -

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Used to identify the external and internal factors influencing the system.
- Used to show the interaction among the requirements are actors.

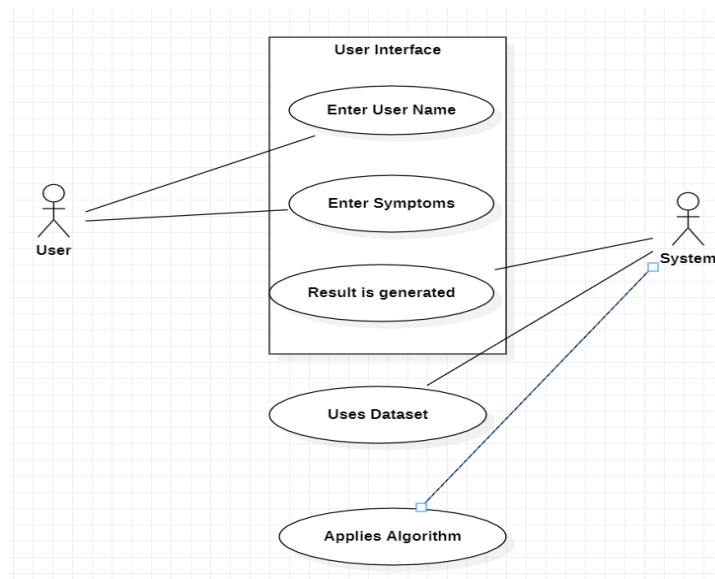


Fig: 4.1.1 Use Case Diagram

4.1.2 CLASS DIAGRAM

Self diagnosable human disease prediction using machine learning (ML) consists of class diagram that all the other application that consist the basic class diagram, here the class diagram is the basic entity that is required in order to carry on with the project. Class diagram consist the data about all the classes that is used and all the related datasets, and all the other necessary attributes and their relationships with other entities, all these information is necessary in order to use the concept of the prediction, where the user will enter all necessary information that is required in order to use the system.

PURPOSE OF CLASS DIAGRAMS

The purpose of a class diagram can be described as -

- Analysis and design of the static view of an application.
- Describe responsibilities of a system.
- Base for component and deployment diagrams.
- Forward and reverse engineering

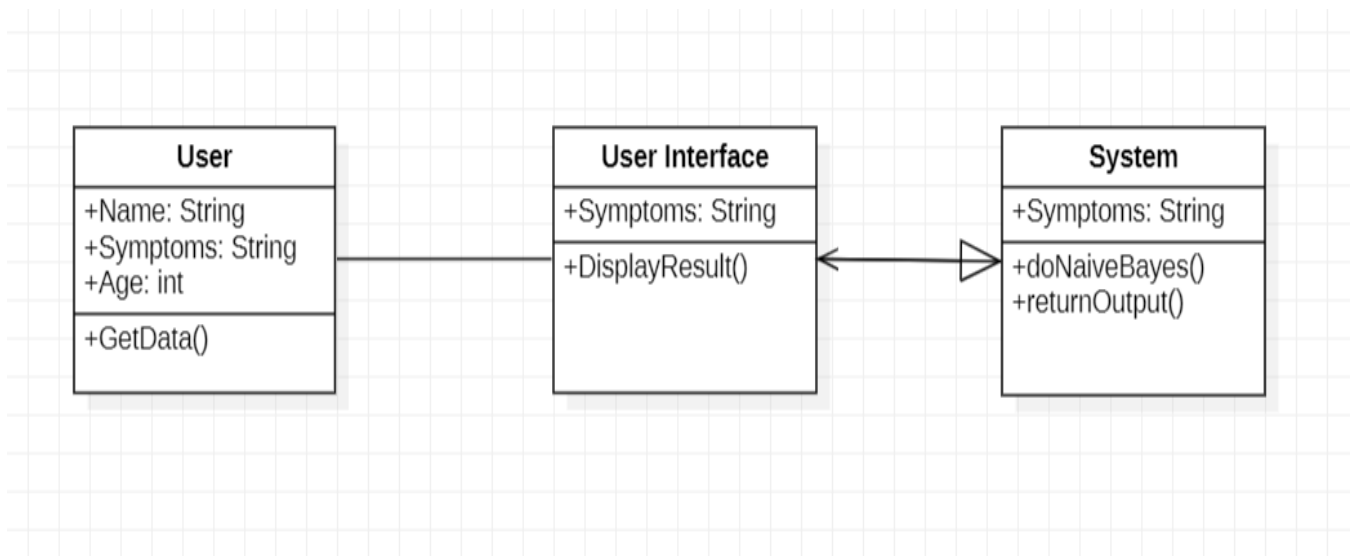


Fig: 4.1.2 Class Diagram

4.1.3 ACTIVITY DIAGRAM

Activity diagram is an important diagram in UML to describe the dynamic aspects of any system. Activity diagram is a flowchart to show the flow from one activity to another. The activity can be explained as an operation of the system's execution. The control flow is taken from one operation to another operation. Here in this diagram the activity starts from user then the user proceeds to the prediction phase where the prediction happens. Then finally after processing the data from datasets the analysis will happen then the correct result or prediction will be displayed which is nothing but the Output.

PURPOSE OF ACTIVITY DIAGRAMS

The purpose of an activity diagram can be described as -

- Draw the activity flow of a system.
- Describe the sequence from one activity to another.
- Describe the parallel, branched and concurrent flow of the system.

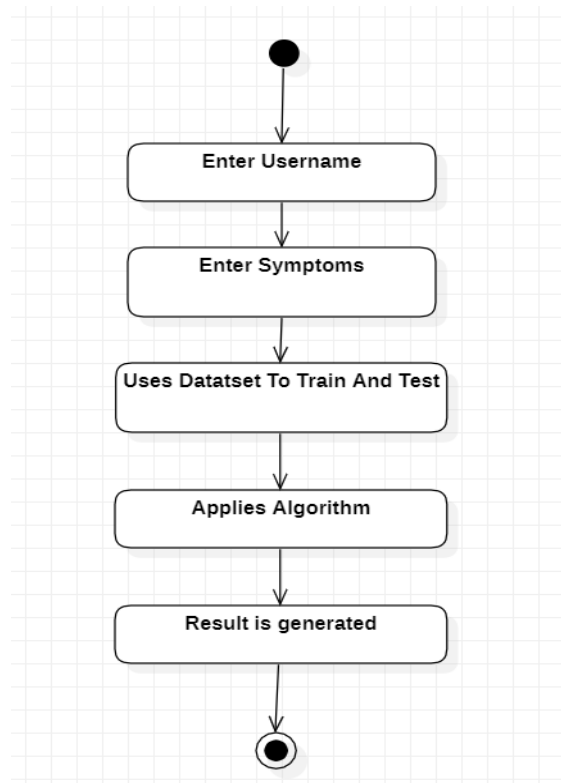


Fig: 4.1.3 Activity Diagram

4.1.4 SEQUENCE DIAGRAM

The Sequence diagram of the project self diagnosable human disease prediction using machine learning (ML) consist of all the various aspects a general sequence diagram requires. This sequence diagram shows how from starting the model flows from one step to another, like how a user enter into the system then enters all the information like symptoms that goes into the system, compares with the prediction model and if true is predicts the appropriate output will be shown otherwise it shows the details where the user gone wrong while entering the information and it also shows the appropriate precautionary measure for the user to follow. Here, the sequence of the entities are linked to each other where the user gets started with the system.

PURPOSE OF SEQUENCE DIAGRAM

- To model the flow of control by time sequence.
- To model the flow of control by structural organizations.
- For forward engineering.
- For reverse engineering.

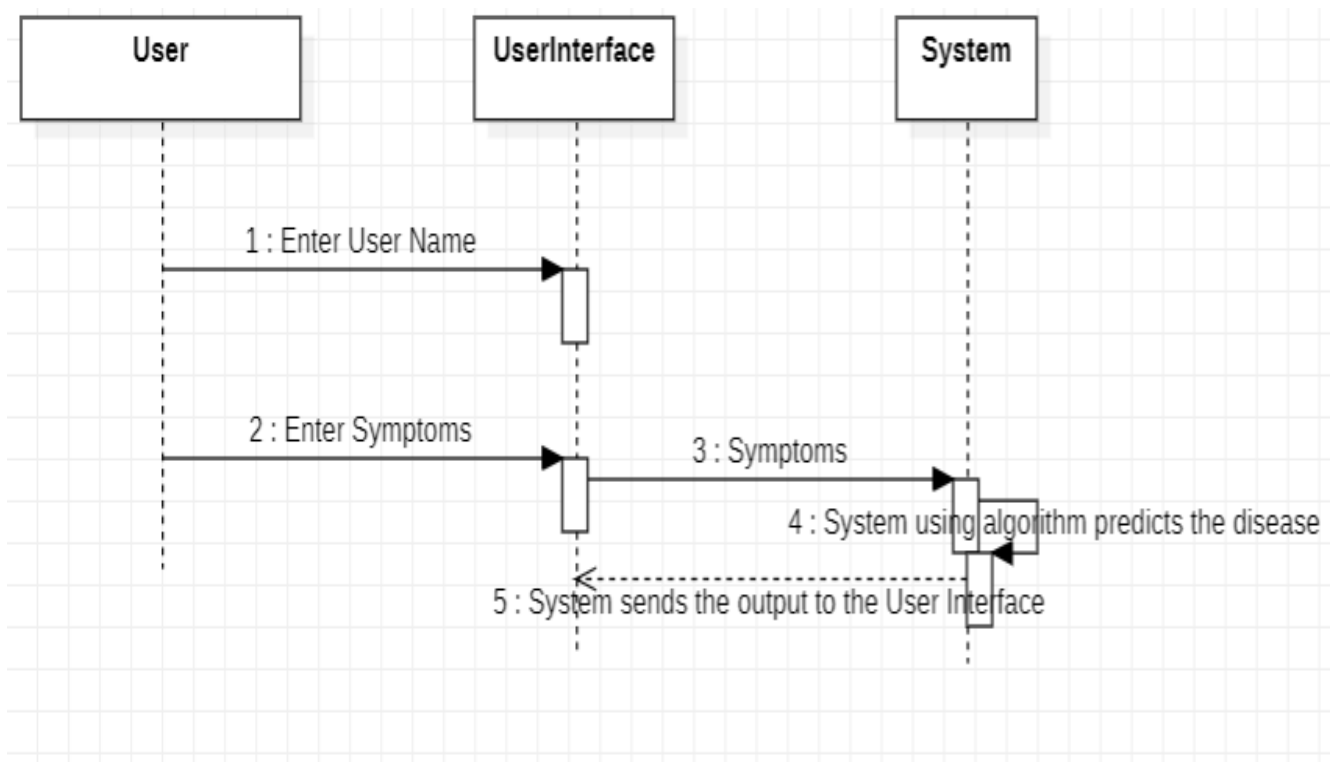


Fig: 4.1.4 Activity Diagram

5. IMPLEMENTATION

5.1 MODULES

A module is a collection of source files and build configurations that allow you to divide your project into many units of functionality. Your project can have one many modules and one module may use another module as a dependency. Each module can be independently built, tested, and debugged A module in project open is a high-level description of a functional area, consisting of a group of processes describing the functionality of the module and a group of packages implementing the functionality We have Four modules in our project namely:

1. User Module
2. Pre-processing Module
3. Training Module
4. Detection Module
5. Testing Module

5.2 MODULE DESCRIPTION

5.2.1 USER MODULE

In User module, the user gives the input as symptoms to the system and the system pre-process the given symptoms and produces an output as disease.

5.2.2 PRE-PROCESSING MODULE

In this module, the symptoms are pre-processed to extract the useful features from the given image and it also does various functions like image adjustment, pixel changing.

5.2.3 TRAINING MODULE

Once the model is created, it has to be trained. We have predefined dataset which contains the symptoms and diseases. This particular dataset is used to train the model.

5.2.4 DETECTION MODULE

In this module the model predicts the disease of a user. If the user enters minimum of two symptoms or maximum of five symptoms then it returns the disease as an output and if the user enters one symptom or if user doesn't enter then it will display a dialog box.

5.3 INTRODUCTION TO TECHNOLOGIES USED

5.3.1 Python

Python is a multi-paradigm programming language. Object oriented programming and structured programming is fully supported, and many of its features support functional programming and object oriented programming. Many other paradigms are also supported via extensions, including design by contract and logic programming. Python uses dynamic typing method and a combination of reference counting and a cycle detecting garbage collector for memory management. It also features dynamic name resolution which is also called as late binding, which binds method and variable names during program execution i.e., when program is running. Python's developers strive to avoid premature optimization, and reject patches to non-critical parts of Python that would offer marginal increases in speed at the cost of clarity. When speed is primary concern then a Python programmer can move time-critical functions to extension modules written in languages such as C, or use PyPy, a just-in-time compiler which can be used. Cython or CPython is also available, which translates a Python script into C and makes direct C level API calls into the Python interpreter. A primary goal of Python's developers is that keeping the language easy and fun to use. Python's design offers some support for functional programming in the Lisp like tradition. It has filter, map, and reduce functions or methods, list comprehensions, dictionaries, sets and generator expressions. The standard library of python has two modules they are itertools and functools that implement functional tools borrowed from Haskell and Standard ML languages.

BENEFITS OF PYTHON

- Presence of Third-Party Modules
- Extensive Support Libraries
- Open Source and Community Development
- Learning Ease and Support Available
- User-friendly Data Structures
- Productivity and Speed
- Highly Extensible and Easily Readable Language.

5.3.2 TKINTER INTERFACE

Tkinter is a Python's version in binding to the Tk GUI toolkit. It is the standard Python interface to the Tk GUI toolkit and is Python's de facto standard GUI. Tkinter is included with standard Linux, Microsoft Windows and Mac OS X installs of Python. The name Tkinter comes from Tk interface. Tkinter was written by Fredrik Lundh. Tkinter is free and open software released under a Python license.

As with most other modern Tk bindings, Tkinter is implemented as a Python's wrapper around a complete Tool Command Language (TCL) interpreter embedded in the Python interpreter. Tkinter calls are translated into Tcl commands which are given to this embedded interpreter, thus making it possible to mix Python and TCL in a single application. The Frame widget is the basic unit of organization for complex layouts in tkinter. A rectangular area called frame can contain other widgets. When any widget is created, a parent child relationship is established. For example, if you place a text label inside a frame, the frame is the parent of the label and so on.

Python offers many options for developing GUI (Graphical User Interface).

Out of all the GUI functions or methods, tkinter is most commonly used function or method. It is a standard Python interface to the Tk GUI toolkit bundled with Python. The python with tkinter outputs the fastest and easiest way to create the GUI applications very conveniently.

TO CREATE A TKINTER:

Import the module tkinter.

Create the main window (container).

Add any number of widgets to the main window.

Apply the event trigger on the widgets.

Importing tkinter is same as importing any other module in the python code.

Note that the name of the module in Python 2.x is 'Tkinter' and in Python 3.x is 'tkinter'.

5.3.3 MACHINE LEARNING

Machine learning (ML) is the study of computer algorithms that improve automatically through experience and no of times it runs. It is seen as a part of artificial intelligence (AI). Machine learning algorithms build a model based on the sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed by a programmer. Machine learning algorithms are used in many applications, such as email filtering and computer vision, where it is difficult to develop conventional algorithms to perform the needed tasks.

FEATURES OF MACHINE LEARNING

- It is nothing but automating the Automation.
- Getting computers to program themselves.
- Machine learning models involve machines learning from data without the help of humans or any kind of human intervention.
- Machine Learning is the science of making the computers learn and act like humans by feeding data and information without being explicitly programmed.

5.4 SAMPLE CODE

```
from tkinter import *
from tkinter import messagebox
from sklearn.naive_bayes import MultinomialNB
from sklearn.metrics import accuracy_score
import numpy as np
import pandas as pd
import warnings
```

```
l1=['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','swollen_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_extremities','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','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']
```

```
disease = ['Fungal infection','Allergy','GERD','Chronic cholestasis','Drug Reaction','Peptic ulcer disease','AIDS','Diabetes','Gastroenteritis','Bronchial Asthma','Hypertension','Migraine','Cervical spondylosis','Paralysis (brain hemorrhage)','Jaundice','Malaria','Chicken pox','Dengue','Typhoid','hepatitis A','Hepatitis B','Hepatitis C','Hepatitis D','Hepatitis E','Alcoholic hepatitis','Tuberculosis','Common Cold','Pneumonia','Dimorphic hemmorhoids(piles)','Heart attack','Varicose veins','Hypothyroidism','Hyperthyroidism','Hypoglycemia','Osteoarthritis','Arthritis','(vertigo) Parosymal Positional Vertigo','Acne','Urinary tract infection','Psoriasis','Impetigo']
```

```
l2 = []
for x in range(0,len(l1)):
    l2.append(0)
```

```
# Testing Data
```

```
testing = pd.read_csv("Testing.csv")
testing.replace({'prognosis':{'Fungal infection':0,'Allergy':1,'GERD':2,'Chronic cholestasis':3,'Drug Reaction':4,'Peptic ulcer disease':5,'AIDS':6,'Diabetes':7,'Gastroenteritis':8,'Bronchial Asthma':9,'Hypertension':10,'Migraine':11,'Cervical spondylosis':12,'Paralysis (brain hemorrhage)':13,'Jaundice':14,'Malaria':15,'Chicken pox':16,'Dengue':17,'Typhoid':18,'hepatitis A':19,'Hepatitis B':20,'Hepatitis C':21,'Hepatitis D':22,'Hepatitis E':23,'Alcoholic hepatitis':24,'Tuberculosis':25,'Common Cold':26,'Pneumonia':27,'Dimorphic hemmorhoids(piles)':28,'Heart attack':29,'Varicose veins':30,'Hypothyroidism':31,'Hyperthyroidism':32,'Hypoglycemia':33,'Osteoarthritis':34,'Arthritis':35,'(vertigo) Parosymal Positional Vertigo':36,'Acne':37,'Urinary tract infection':38,'Psoriasis':39,'Impetigo':40}},inplace=True)
```

```
X_test = testing[l1]
y_test = testing[["prognosis"]]
```

```

np.ravel(y_test)

# Training Data
training = pd.read_csv("Training.csv")
training.replace({'prognosis':{'Fungal infection':0,'Allergy':1,'GERD':2,'Chronic
cholestasis':3,'Drug Reaction':4,
'Peptic ulcer disease':5,'AIDS':6,'Diabetes ':7,'Gastroenteritis':8,'Bronchial
Asthma':9,'Hypertension ':10,
'Migraine':11,'Cervical spondylosis':12,
'Paralysis (brain hemorrhage)':13,'Jaundice':14,'Malaria':15,'Chicken
pox':16,'Dengue':17,'Typhoid':18,'hepatitis A':19,
'Hepatitis B':20,'Hepatitis C':21,'Hepatitis D':22,'Hepatitis E':23,'Alcoholic
hepatitis':24,'Tuberculosis':25,
'Common Cold':26,'Pneumonia':27,'Dimorphic hemmorhoids(piles)':28,'Heart
attack':29,'Varicose veins':30,'Hypothyroidism':31,
'Hyperthyroidism':32,'Hypoglycemia':33,'Osteoarthritis':34,'Arthritis':35,
'(vertigo) Paroymsal Positional Vertigo':36,'Acne':37,'Urinary tract
infection':38,'Psoriasis':39,
'Impetigo':40}} ,inplace=True)

X = training[1:]

y = training[["prognosis"]]
np.ravel(y)

# Methods

def message():
    if (Symptom1.get() == "None" and Symptom2.get() == "None" and Symptom3.get() ==
"None" and Symptom4.get() == "None" and Symptom5.get() == "None"):
        messagebox.showinfo("Oops!!", "Enter some symptoms please")
    else :
        NaiveBayes()

def doNaiveBayes():
    gnb = MultinomialNB()
    gnb = gnb.fit(X,np.ravel(y))
    from sklearn.metrics import accuracy_score
    y_pred = gnb.predict(X_test)

```



```

print("Accuracy:", accuracy_score(y_test, y_pred))
print("Accuracy Normalized:", accuracy_score(y_test, y_pred, normalize=False))
psymptoms=[Symptom1.get(),Symptom2.get(),Symptom3.get(),Symptom4.get(),Symptom5
.get()]

for k in range(0,len(l1)):
    for z in psymptoms:
        if(z == l1[k]):
            l2[k] = 1

inputtest = [l2]
predict = gnb.predict(inputtest)
predicted=predict[0]

h='no'
for a in range(0,len(disease)):
    if(disease[predicted] == disease[a]):
        h='yes'
        break

if (h=='yes'):
    displayResult (disease[a])
else:
    displayResult ("No Disease")

def displayResult(msg):
    display.configure(state=NORMAL)
    display.delete("1.0", END)
    display.insert(END, msg)
    display.configure(state=DISABLED)

def getData():
    name = input("Enter your name: ")
    age = input("Enter your age: ")
    p = open("save.txt", "w")
    p.write("User Name: " + name + "\nUser Age: " + age)
    p.close()

# Main Window
window = Tk()
window.title("Self Diagnosable Human Disease Prediction")

```

```

window.configure()
warnings.simplefilter("ignore")
getData()

# Variables that store option menu selected value
Symptom1 = StringVar()
Symptom1.set(None)
Symptom2 = StringVar()
Symptom2.set(None)
Symptom3 = StringVar()
Symptom3.set(None)
Symptom4 = StringVar()
Symptom4.set(None)
Symptom5 = StringVar()
Symptom5.set(None)

# Area to store the components in a grid
root = Frame(window)
root.pack()

# Project heading
heading = Label(root, justify=LEFT, text="Self Diagnosable Human Disease Prediction")
heading.config(font=("Times New Roman", 26))
heading.grid(row=1, column=1, columnspan=2, padx=10)

# Symptom labels

NameLb1 = Label(root, text="")
NameLb1.config(font=("Times New Roman", 20))
NameLb1.grid(row=5, column=1, pady=10, sticky=W)

S1Lb = Label(root, text="Symptom 1")
S1Lb.config(font=("Times New Roman", 15))
S1Lb.grid(row=7, column=1, padx=10, pady=10, sticky=W)

S2Lb = Label(root, text="Symptom 2")
S2Lb.config(font=("Times New Roman", 15))
S2Lb.grid(row=8, column=1, padx=10, pady=10, sticky=W)

S3Lb = Label(root, text="Symptom 3")

```

```

S3Lb.config(font=("Times New Roman", 15))
S3Lb.grid(row=9, column=1, padx=10, pady=10, sticky=W)

S4Lb = Label(root, text="Symptom 4")
S4Lb.config(font=("Times New Roman", 15))
S4Lb.grid(row=10, column=1, padx=10, pady=10, sticky=W)

S5Lb = Label(root, text="Symptom 5")
S5Lb.config(font=("Times New Roman", 15))
S5Lb.grid(row=11, column=1, padx=10, pady=10, sticky=W)

# Option Menu
OPTIONS = sorted(l1)

S1En = OptionMenu(root, Symptom1,*OPTIONS)
S1En.grid(row=7, column=2, padx=10)

S2En = OptionMenu(root, Symptom2,*OPTIONS)
S2En.grid(row=8, column=2, padx=10)

S3En = OptionMenu(root, Symptom3,*OPTIONS)
S3En.grid(row=9, column=2, padx=10)

S4En = OptionMenu(root, Symptom4,*OPTIONS)
S4En.grid(row=10, column=2, padx=10)

S5En = OptionMenu(root, Symptom5,*OPTIONS)
S5En.grid(row=11, column=2, padx=10)

# Predict Button
predict_btn = Button(window, text="Predict",height=2, width=20, command=message)
predict_btn.config(font=("Times New Roman", 15))
predict_btn.pack(padx=10, pady=20)

# Display Box
display = Text(window, height=2, width=30)
display.config(font=("Times New Roman", 20))
display.pack(padx=10, pady=10)
display.configure(state=DISABLED)
root.mainloop()

```

6. TEST CASES

This system can predict disease based on symptoms.

INPUT:

Symptom 1: cough

Symptom 2: constipation

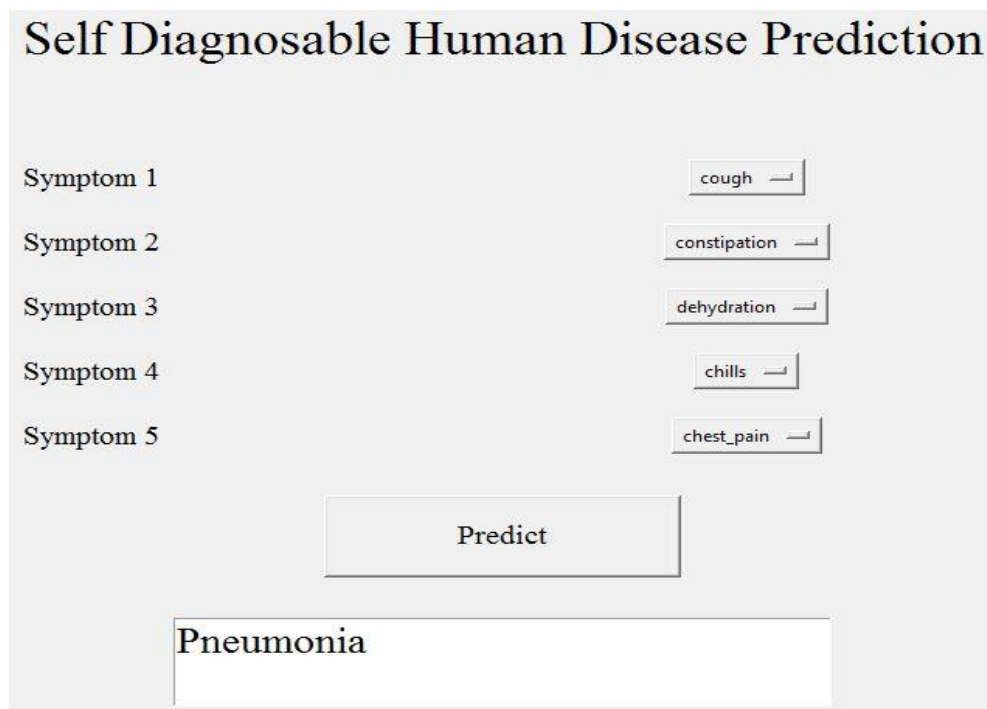
Symptom 3: dehydration

Symptom 4: chills

Symptom 5: chest pain

OUTPUT:

Pneumonia



Self Diagnosable Human Disease Prediction

Symptom 1

Symptom 2

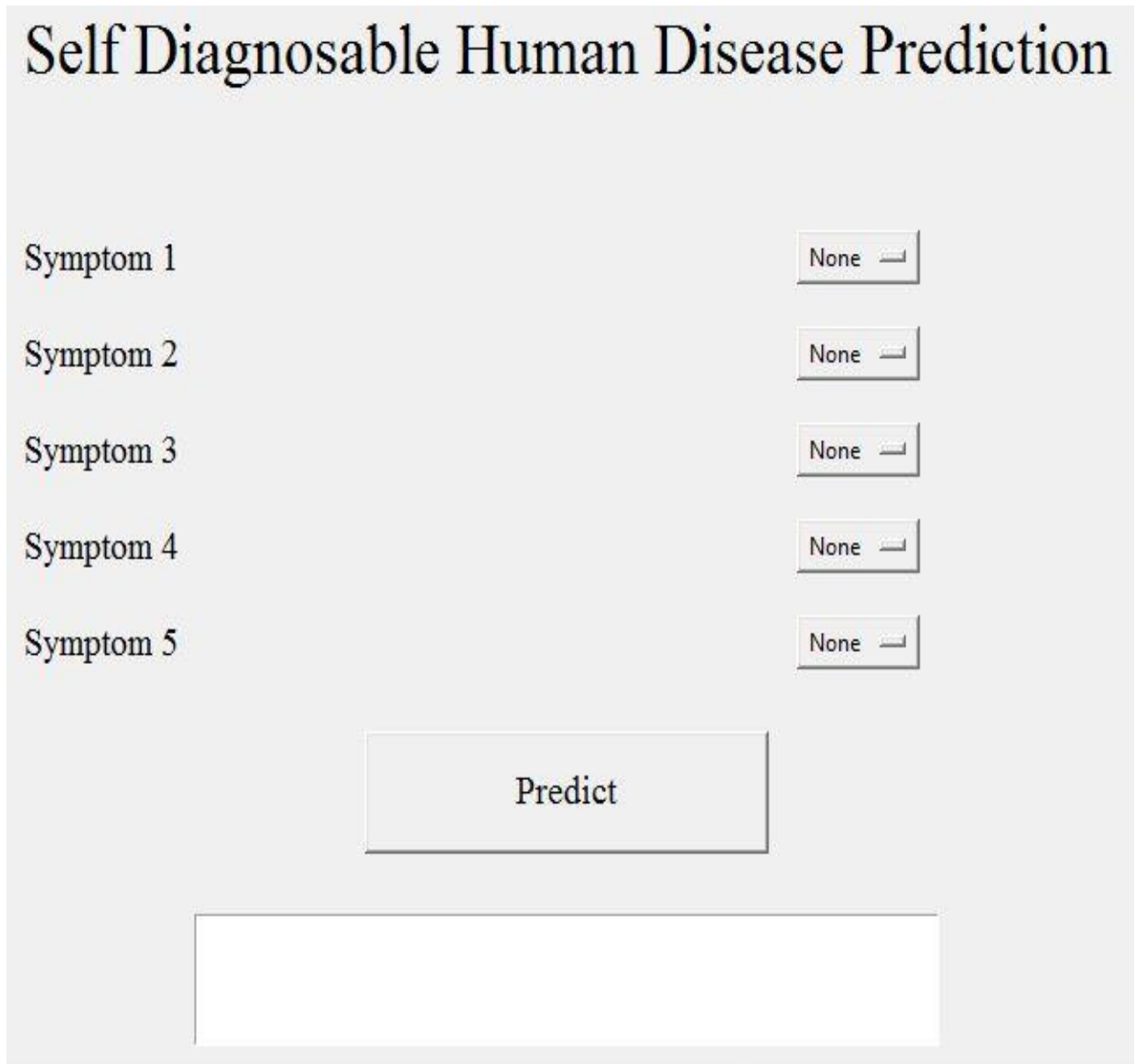
Symptom 3

Symptom 4

Symptom 5

Fig. 6.1: Extracted image of output

7. SCREENSHOTS



The screenshot shows a web form titled "Self Diagnosable Human Disease Prediction". It contains five rows, each with a symptom label and a dropdown menu. All dropdown menus are currently set to "None". Below the symptom rows is a "Predict" button, and at the bottom is a large empty rectangular box for the prediction result.

Symptom	Value
Symptom 1	None
Symptom 2	None
Symptom 3	None
Symptom 4	None
Symptom 5	None

Predict

Fig: 7.1

Self Diagnosable Human Disease Prediction

Symptom 1	<input type="text" value="cough"/>
Symptom 2	<input type="text" value="congestion"/>
Symptom 3	<input type="text" value="None"/>
Symptom 4	<input type="text" value="None"/>
Symptom 5	<input type="text" value="None"/>

Fig: 7.2

Self Diagnosable Human Disease Prediction

Symptom 1	<input type="text" value="bloody_stool"/>
Symptom 2	<input type="text" value="belly_pain"/>
Symptom 3	<input type="text" value="diarrhoea"/>
Symptom 4	<input type="text" value="None"/>
Symptom 5	<input type="text" value="None"/>

Fig: 7.3

8. CONCLUSION

So, finally I conclude by describing that, this project self diagnosable human disease prediction using machine learning (ML) is very useful in everyone's everyday life and it is mainly more important for the healthcare sector, because they are the one that daily uses these systems to predict the diseases of the patients based on their general information and their symptoms that they have been through. Now-a-day's health industry plays major role in treating the diseases of the patients so this is also some kind of help for the health sector to tell the user and also it is much useful for the user in case he/she doesn't want to go to the hospital then they can just select the symptoms and the user can get to know the disease that they are suffering from and the health industry can also get benefited from this system by just asking the symptoms from the user and entering in the system and then in just a few seconds they can tell the exact and up to some extent the accurate diseases. If health industry takes up this project then the work of the doctors can be reduced drastically and they can easily predict the disease of the patient very quickly. The Self Diagnosable Human Disease prediction project is to provide prediction for the various generally occurring diseases that when unchecked and sometimes ignored can turn into dangerous disease and cause a lot of problem to the patient and as well as the people in their family. So, many people who do not want to go to hospitals and clinics they can use this project to know their health and disease condition very conveniently and quickly sitting in their homes.

9. FUTURE ENHANCEMENTS

As now-a-days we can clearly witness the increase in use of computers and technology to consider a huge amount of data, computers are being used to perform various complex tasks with commendable accuracy rates. Machine learning (ML) is a collection of multiple techniques and algorithms which permit computers to execute such complex tasks in a simplified manner. It is also used in both academics which is for students or learners and also in industry to make accurate predictions and use these diverse sources of dataset and information. Till this date we can say we got developed in the fields of Big Data, Machine learning, and Data Sciences etc and have been a part of one of those big industries which were able to collect such data and the staff to transform their goods and services in a desired manner. The learning methods developed for these industries and researches offer excellent potential to further improve in medical research and clinical care for the patients in the best possible manner. Machine learning uses mathematical algorithms and procedures which are used to describe the relationship between variables used in the model and the others. Our paper will explain the process of training the model and learning a suitable algorithm to predict the presence of a particular disease from the sample of the tissue based on its features. Though these algorithms work in different and unique manners depending on the way in which they are developed and used by the researchers. One method is to consider their main goals. The goal of our project is to reach to a conclusion about the data which are collected from a wide variety of samples from our population. Though many techniques, like bayes theorem can be able to predict the diseases. For example, consider a case where, if we can create a model which described and understood the relationship between clinical variables and their transience then we can follow the organ transplant surgery i.e. we would need the factors and features which differentiate low mortality rate from high if we can develop such outcomes and reduce mortality rate to a desired rate in the future and also nothing can be said to be better than such situations.

10. BIBLIOGRAPHY

1. D. W. Bates, S . Saria, L. Ohno-Machado, A. Shah, and G. Escobar, “Big data in health care: using analytics to identify and manage high-risk and high-cost patients,” Health Affairs.
2. K.R.Lakshmi, Y.Nagesh and M.VeeraKrishna, ”Performance comparison of three data mining techniques for predicting kidney disease survivability”, International Journal of Advances in Engineering Technology.
3. Mr. Chala Beyene, Prof. Pooja Kamat, “Survey on Prediction and Analysis the Occurrence of Heart Disease Using Data Mining Techniques”, International Journal of Pure and Applied Mathematics.
4. Boshra Brahmi, Mirsaeid Hosseini Shirvani, “Prediction and Diagnosis of Heart Disease by Data Mining Techniques”, Journals of Multidisciplinary Engineering Science and Technology.