DISEASE PREDICTION USING MACHINE LEARNING

1. INTRODUCTION

1.1 Project Overview

In today's rapidly evolving world, where the constraints of time often lead to the neglect of one's health, we find a pressing need for innovative solutions that prioritize wellness and medical care. With the bustling pace of modern life, individuals frequently put off seeking medical attention, even when they experience severe or perplexing symptoms of various diseases. The reliance on a simple Google search for symptom-checking often yields generic or stereotypical results, leaving people frustrated, anxious, and uncertain about their health.

Recognizing the urgency of these challenges, we have introduced a groundbreaking solution that harnesses the power of artificial intelligence and predictive modeling. Our solution is a state-of-the-art predictive model capable of identifying up to 42 different diseases when provided with a list of symptoms. This innovative model has the potential to transform the landscape of healthcare by offering a unique and accessible approach to medical diagnosis.

1.2 Purpose

The primary purpose of this project is to address the critical issues of healthcare accessibility, timely diagnosis, and individual empowerment. Our groundbreaking predictive model is designed to serve a multifaceted purpose:

1. **Revolutionizing Medical Diagnosis:** At its core, our project aims to revolutionize the way diseases are diagnosed and medical advice is sought. By utilizing advanced predictive modeling, we aim to bridge the gap between the onset of symptoms and the initiation of medical treatment. This approach is particularly crucial for conditions that demand immediate attention.
2. **Enhancing Telemedicine and Online Consultations:** Healthcare professionals face the challenge of providing accurate diagnoses and consultations, especially in remote or telemedicine settings. Our model equips them with a powerful diagnostic aid that complements their expertise, enabling them to consult with patients remotely based on the model's predictions. This capability is a game-changer for telemedicine services and online consultations, allowing for more accurate and timely medical advice.

3. **Empowering Individuals for Preventive Care:** The high costs associated with visiting a doctor and the inconvenience of scheduling appointments can be significant barriers to seeking medical advice for many individuals. Our model, however, empowers individuals to take control of their health and make informed decisions without the need to share personal data such as their name, age, gender, or address. It promotes preventive care by enabling users to assess their symptoms and make informed choices about their healthcare.
4. **User-Friendly Access:** We have developed a user-friendly web application that ensures that anyone, regardless of their technological proficiency, can access our model with ease. It provides users with a straightforward interface for entering symptoms and receiving probable disease identifications, promoting a seamless and accessible healthcare experience.

2. LITERATURE SURVEY

2.1 Existing problem

In the realm of healthcare and medical diagnostics, there exists a long-standing challenge that pertains to the timely and accurate identification of a wide spectrum of diseases. With the advent of Machine Learning and Deep Learning technologies, there is a substantial opportunity to address this issue. The traditional healthcare system is often inundated with a multitude of patients, and people often find it cumbersome to visit a healthcare professional for every instance of ailment. This delay in seeking medical attention or the reliance on generic internet searches can lead to misdiagnosis, anxiety, and potentially adverse health outcomes. As the world becomes increasingly fast-paced, there is a growing need to streamline the process of disease identification and provide individuals with a more convenient and effective means of obtaining initial assessments of their health. This problem is not only of medical significance but also reflects a larger societal issue: the need for accessible and efficient healthcare.

2.2 References

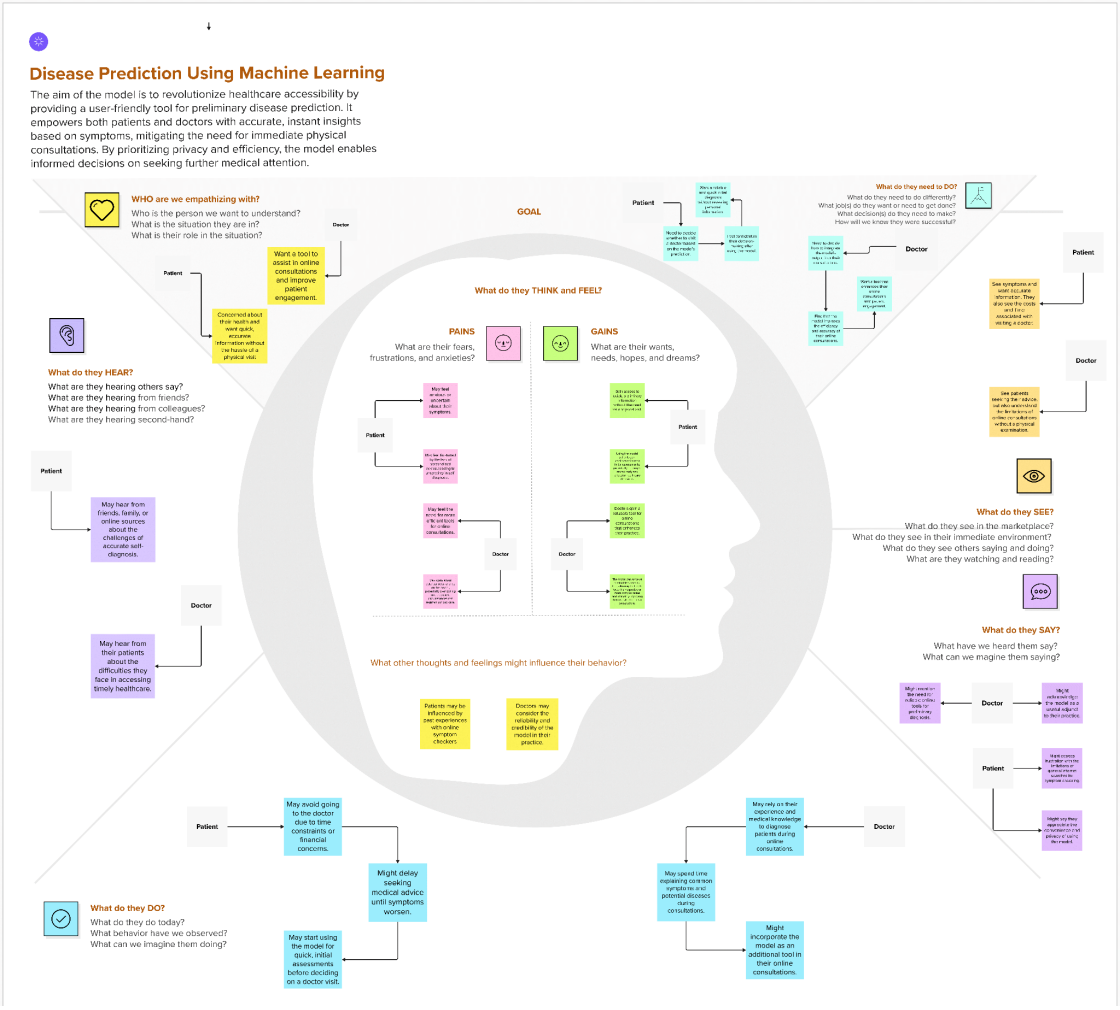
To comprehensively address the existing problem of disease identification, we turn to a wealth of research, technological advancements, and data sources. In our endeavor to classify 42 diseases, we draw upon a multitude of scholarly references and academic studies within the fields of machine learning, deep learning, and healthcare diagnostics. These references serve as the foundation upon which our project is built, allowing us to leverage the collective knowledge and expertise of the global scientific and technological community. By examining and synthesizing the insights and methodologies from these references, we aim to create a powerful and informed solution that advances the state of the art in disease prediction using machine learning.

2.3 Problem Statement Definition

The problem we aim to address can be succinctly defined as follows: We seek to utilize Machine Learning and Deep Learning models to classify 42 different diseases accurately. To achieve this, we intend to build upon the Kaggle dataset for Disease Prediction Using Machine Learning, which provides a substantial repository of medical images and associated symptom data. Our project's objective is to train a model that can predict the likelihood of a disease based on a user-provided list of up to nine symptoms. By doing so, we aim to streamline and democratize the disease identification process, making it more user-centric, efficient, and accessible. This project serves to bridge the gap between traditional healthcare systems and modern technology, empowering individuals to make informed decisions about their health while contributing to the ongoing evolution of healthcare diagnostics through advanced machine learning techniques.

3. IDEATION & PROPOSED SOLUTION

3.1 Empathy Map Canvas

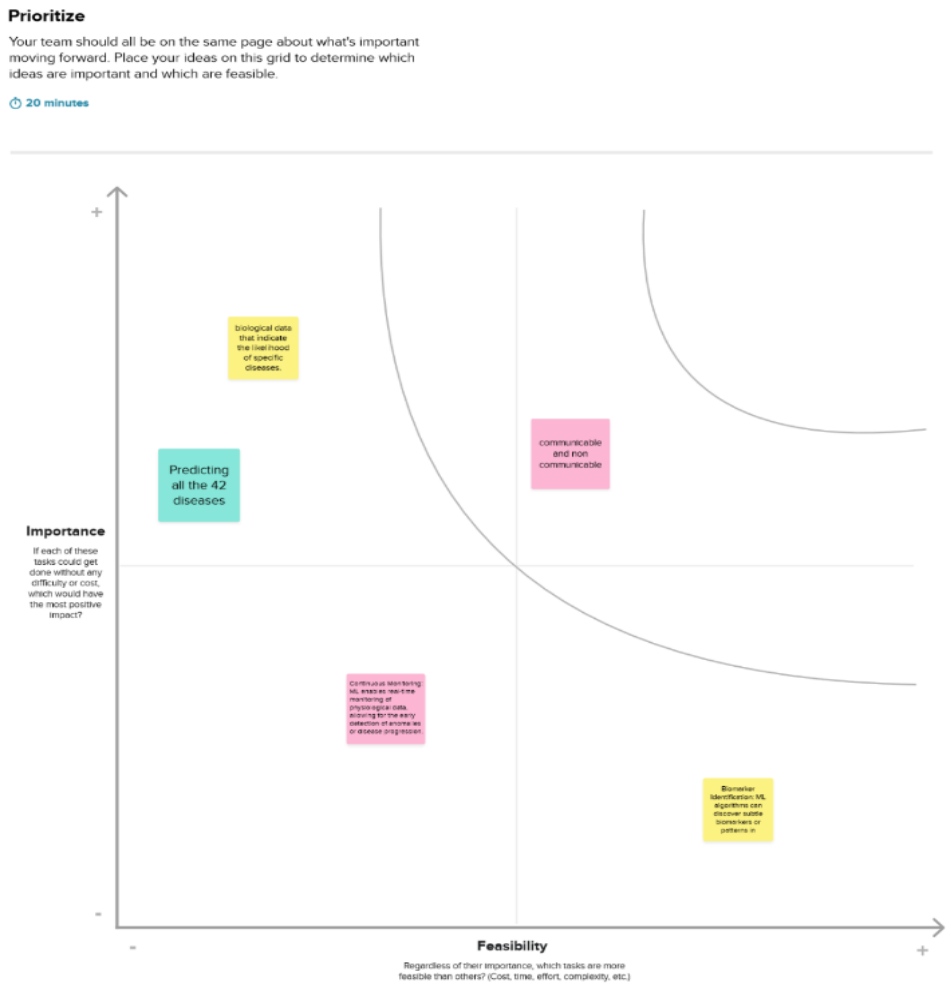


https://app.mural.co/t/smartinternz5563/m/smartinternz5563/1697545979100/04105f1ec0d4cc42a7d24a81f8d465962fb92eed?sender=u8c1b28ad684e4f6a2ad86663

3.2 Ideation & Brainstorming

Brainstorming provides a free and open environment that encourages everyone within a team to participate in the creative thinking process that leads to problem solving. Prioritizing volume over value, out-of-the-box ideas are welcome and built upon, and all participants are encouraged to collaborate, helping each other develop a rich amount of creative solutions. Use this template in your own brainstorming sessions so your team can unleash their imagination and start shaping concepts even if you're not sitting in the same room.


IDEAS
Implement real-time symptom monitoring through wearable devices for continuous health tracking
Consider creating a user community for sharing health-related experiences and tips, fostering user engagement
Explore the use of AI-driven chatbots to provide immediate responses to user inquiries and guide them through the prediction process.
Investigate potential integration with health insurance companies to offer discounts for preventive healthcare measures.
Different diseases we consider
skin dieseases
heart diseases
viral diseases
fever, cold, allergies.
communicable and non communicable
Personal info we do not ask
name , age, address
phone number email-Id gender religion
age we need to rethink
any legal documents
Features
Continuous Monitoring: ML enables real-time monitoring of physiological data, allowing for the early detection of anomalies or disease progression.
Feature Selection: ML models can automatically identify and prioritize the most relevant factors from a large dataset to predict diseases.
Patient History Analysis: ML can analyze a patient's medical history, including past illnesses, treatments, and family history, to assess disease risk.
User Interface & Accessibility
Create a user-friendly web interface for symptom input.
Implement multilingual support for broader user access.
Design a mobile app version for symptom checks on-the-go.
Include a symptom severity scale for improved accuracy.
color palette: blue,white
Data Integration & Model Enhanceme...
Explore using anonymized patient data to improve model accuracy.
Consider partnerships with research institutions for ongoing model improvement.
Investigate the use of data from multiple sources, such as electronic health records.
Research the potential integration of telehealth features.
Privacy & Security
Ensure compliance with data protection regulations.
Examine privacy and security measures for sensitive health data.
Explore ways to protect against data breaches.
Investigate encryption and secure data storage solutions.
Model Transparency & Education
Investigate the use of explainable AI to make model decisions transparent.
Create a knowledge base to provide users with educational resources.
Develop strategies to educate users about the model's limitations.
Explore options for expanding the model's capabilities to predict outbreaks of infectious diseases.
Marketing & User Acquisition
Think about marketing and user acquisition strategies for the web app
Consider social media integration for sharing health tips and updates
Explore ways to promote the use of the tool for preventive healthcare
Investigate options for integrating with electronic health records (EHR) systems
Business Model & Monetization
Research potential monetization models (e.g., subscription fees or ads)
Consider offering both free and premium versions with added feature
Think about partnerships with health insurance providers
Explore the use of gamification for encouraging regular health check-ins
FINALIZED IDEAS
Biomarker Identification: ML algorithms can discover subtle biomarkers or patterns in
communicable and non communicable
Continuous Monitoring: ML enables real-time monitoring of physiological data, allowing for the early detection of anomalies or disease progression.
biological data that indicate the likelihood of specific diseases.



https://app.mural.co/t/diseasepred9904/m/diseasepred9904/1697549249995/d6866edc6f91e809be518f241932bd41b6cea9ba?sender=u8244eca73f67a84ae5860534

4. REQUIREMENT ANALYSIS

4.1 Functional requirement -

1. **User Input**: - The web application should allow users to input their symptoms. Users can enter multiple symptoms to provide a comprehensive description of their condition
2. **Disease Prediction**: - The system should be able to predict up to 42 diseases based on the symptoms provided by the user.

3. **User Output**: The web application should display the list of probable diseases along with their respective likelihood scores or probabilities.
4. **Self-Care Information**: The system should provide users with general information about the predicted diseases, including common symptoms, treatment options, and when to seek professional medical help.
5. **User Privacy**: The model should not ask for or store any personally identifiable information (PII), such as name, age, gender, religion, address, or other personal details.
6. **Accessibility**:The web application should be designed to be accessible to individuals with disabilities, complying with accessibility standards (e.g., WCAG).

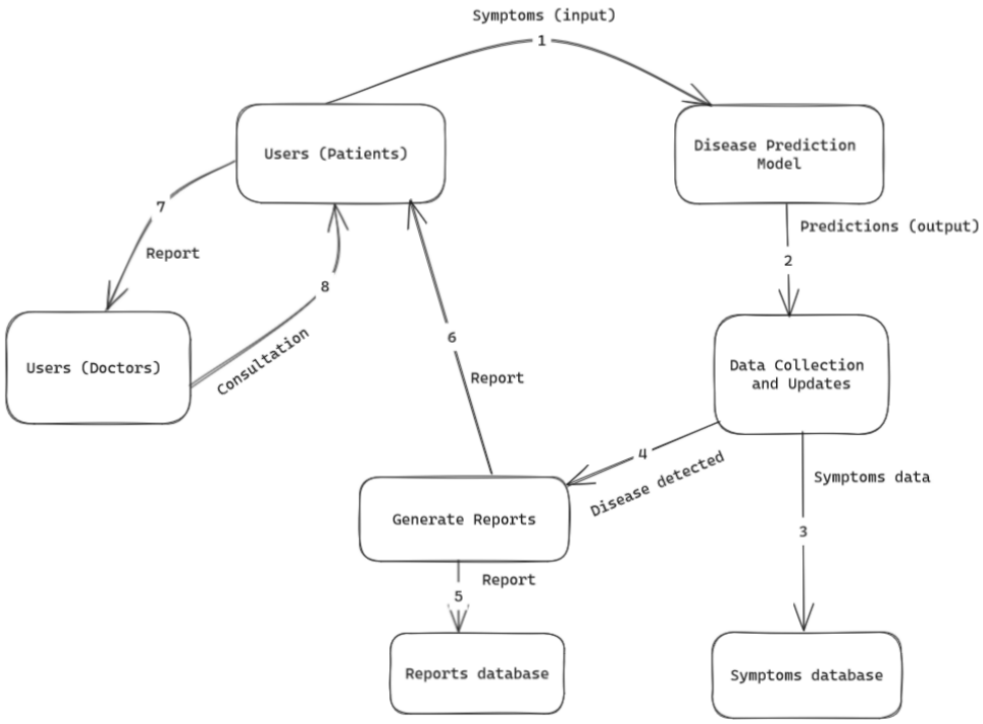
4.2 Non-Functional requirements

1. **Accuracy**: The disease prediction model should be highly accurate in providing likely disease diagnoses based on symptoms. The model should be regularly updated to improve accuracy.
2. **Speed**: The system should provide quick results, ensuring minimal wait times for users. Response times should be optimized to provide a seamless user experience
3. **Security**: Ensure data security and privacy by implementing encryption for data transmission, secure authentication for doctors, and robust access control mechanisms.
4. **Reliability:** The system should be available and reliable 24/7 to cater to users' needs, with minimal downtime for maintenance.
5. **Scalability**: The system should be able to handle a growing number of users and concurrent consultations, without compromising performance.
6. **User-Friendly Interface**: The web application should have an intuitive and user-friendly interface that is easy for users to navigate and interact with.
7. **Maintenance**:Regular maintenance and updates should be carried out to ensure the system's performance, accuracy, and security.
8. **Compliance**: - The system should comply with relevant healthcare and medical data protection regulations and standards, such as HIPAA (if applicable).
9. **Language Support**: - The application should support multiple languages to cater to a diverse user base.
10. **Cost-Effective**:The development and operation of the system should be cost-effective to encourage its use as an alternative to in-person doctor visits.

11. **Feedback Mechanism**: Implement a feedback mechanism to allow users to report any inaccuracies or issues with the predictions, helping to improve the model and system over time.
12. **Data Handling**: Ensure that the system follows best practices for data handling, including anonymizing and securing user-provided symptom data. Data should be stored and transmitted securely.
13. **Compatibility:-** The web application should be compatible with various browsers and devices, including mobile phones and tablets.
14. **Regular Updates**: - The disease prediction model and medical information database should be regularly updated to reflect the latest medical knowledge and research.
15. **Support and Training:** Provide support and training to doctors and users for effectively using the system for online consultations and self-care.

5. PROJECT DESIGN

5.1 Data Flow Diagrams & User Stories



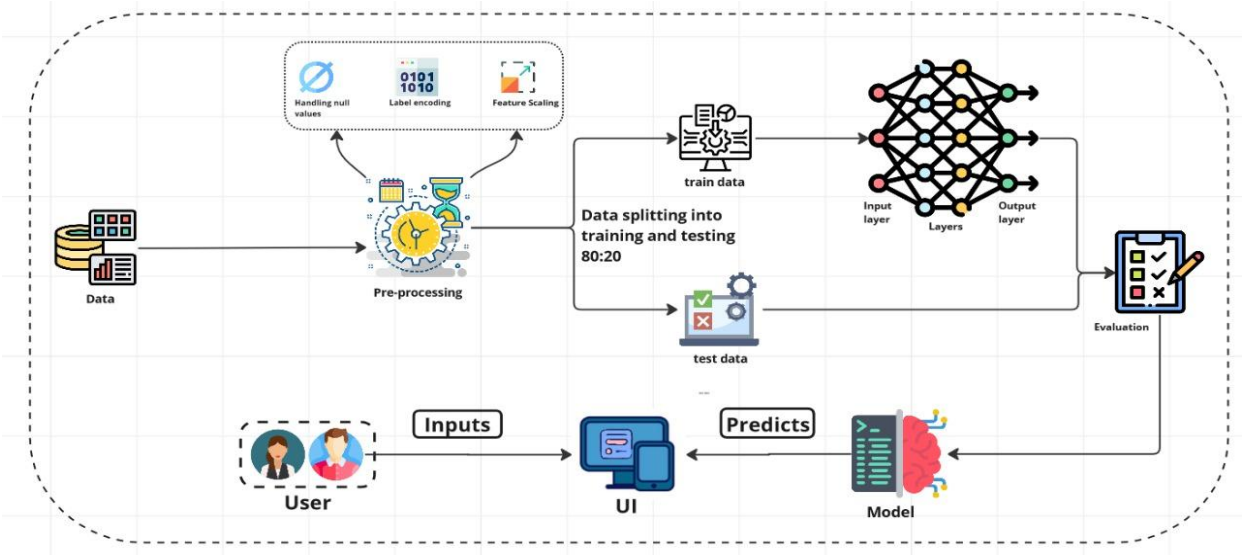
1. **Users (Patients):** Patients input symptoms for the disease prediction model.

2. **Disease Prediction Model:** The core of the system, which predicts diseases based on the symptoms.
3. **Data Collection and Updates:** Stores historical symptom data in the symptoms database for training and updating the model.
4. **Reporting:** Generates reports based on predictions of disease(s) detected.
5. **Reports storage:** Stores the generated reports in the Reports database.
6. **Display Report:** The user (patient) can access their report for preventive diagnosis and self care.
7. **Users (Doctors):** Patients can choose to share their report with a Doctor for consultation.
8. **Consultation:** Online consultations between doctors and patients.

User Stories

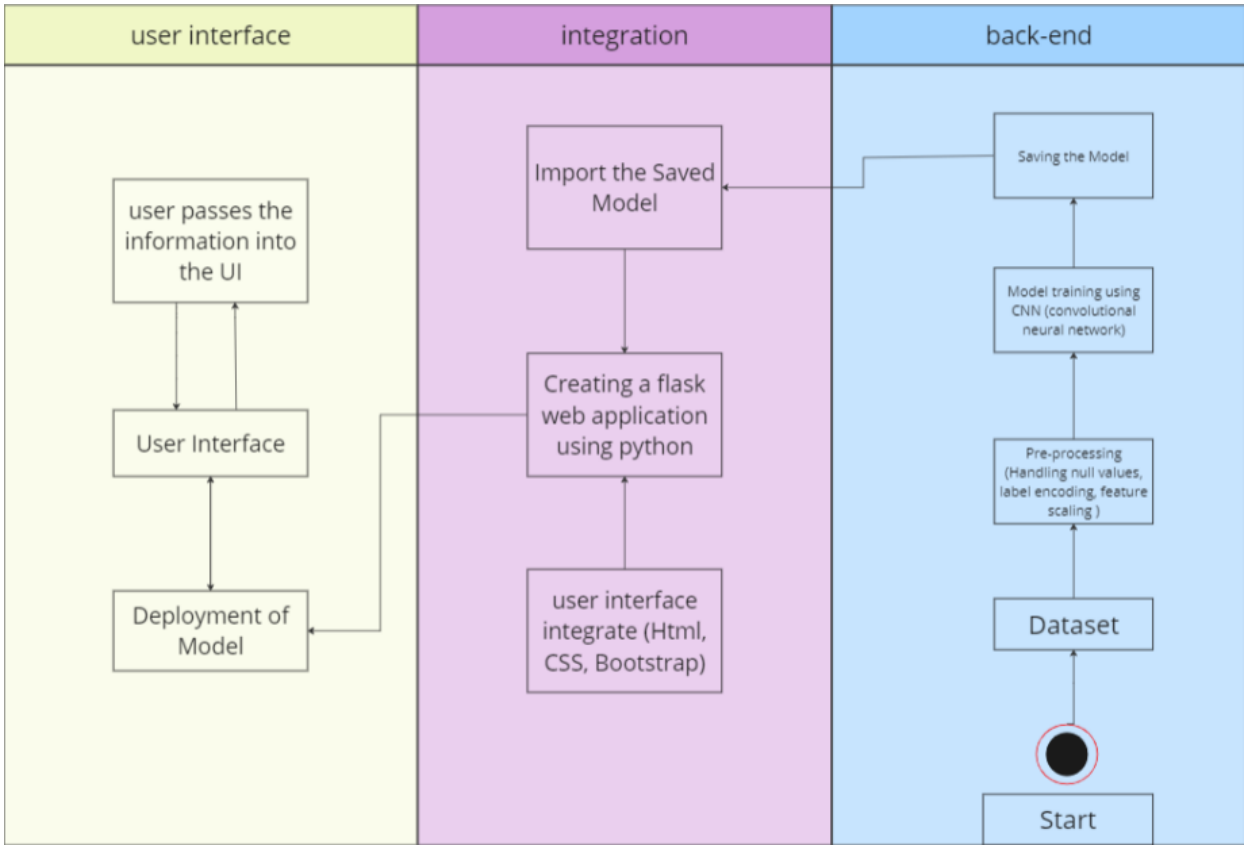
User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Patient	Quick Symptom Check	USN - 1	Input symptoms, receive disease prediction	Users can input symptoms without providing personal information.	High	Sprint - 1
				The model provides a probable disease based on the input symptoms.		
				The system is accessible 24x7 for symptom checks.		
				Users can easily decide whether to visit a doctor based on the prediction.		
Healthcare Professionals	Efficient online consultations	USN - 2	Utilize the model for online consultations	The model's predictions assist in providing more informed advice to patients.	High	Sprint - 1
				Online consultations are conducted efficiently.		
				The model provides a list of possible diseases, not a definitive diagnosis.		
Developers / Administrators	User-friendly and secure application	USN - 3	Develop and maintain the web application	The application is easy to use and navigate.	Medium	Sprint - 2
				The application is secure and protects user privacy.		
				Regular monitoring and updates are in place to ensure accuracy.		
All Users	Accurate and transparent predictions	USN - 4	Input symptoms and receive explanations	The application provides clear explanations of the likelihood of each disease.	High	Sprint - 2
				The application suggests seeking professional medical attention for certain predictions.		
				No personal information is collected or required during symptom input.		
	Easily accessible on various devices	USN - 5	Use the application on different devices	The application is responsive and works well on various devices (PC, mobile, tablet).	Low	Sprint - 3
				Users can access the application from anywhere with an Internet connection.		
	Consult with Healthcare professionals	USN - 6	Seek a professional consultation	Users have the option to contact a healthcare professional for more personalized guidance.	High	Sprint - 2

5.2 Solution Architecture



6. PROJECT PLANNING & SCHEDULING

6.1 Technical Architecture



6.2 Sprint Planning & Estimation

Sprint	Functional Requirement (Epic)	User Story Number	User Story / Task	Story Points	Priority	Team Members
Sprint - 1	Quick Symptom Check	USN - 1	Input symptoms, receive disease prediction	2	High	Mathew
Sprint - 1	Efficient online consultations	USN - 2	Utilize the model for online consultations	1	High	Ryan
Sprint - 2	User-friendly and secure application	USN - 3	Develop and maintain the web application	2	Medium	Priyanshu
Sprint - 2	Accurate and transparent predictions	USN - 4	Input symptoms and receive explanations	2	High	Mathew
Sprint - 3	Easily accessible on various devices	USN - 5	Use the application on different devices	1	Low	Ayushi
Sprint - 2	Consult with Healthcare professionals	USN - 6	Seek a professional consultation	1	High	Ryan

6.3 Sprint Delivery Schedule

Sprint	Total Story Points	Duration (Days)	Sprint Start Date	Sprint End Date (Planned)	Story Points Completed (as on Planned End Date)	Sprint Release Date (Actual)
Sprint - 1	10	8	10.10.2023	18.10.2023	10	17.10.2023
Sprint - 2	10	5	19.10.2023	23.10.2023	10	23.10.2023
Sprint - 3	5	4	24.10.2023	27.10.2023	5	24.10.2023
Sprint - 4	15	10	28.10.2023	06.11.2023		
Sprint - 5	15	3	07.11.2023	09.11.2023		

7. CODING & SOLUTIONING (Explain the features added in the project along with code)

7.1 Feature 1

Symptom-Based Disease Prediction: The model can predict up to 42 diseases based on the symptoms provided as input. This can help users get a general idea of what might be causing their symptoms, potentially saving them time and worry. The web application allows easy access to the disease prediction model, making it available to anyone with an internet connection. Users can access it at any time, making it convenient for self-diagnosis or seeking initial information.

Also, doesn't require users to provide any personalized data like name, age, gender, or address. This focus on privacy can help users feel more comfortable using the service, as their sensitive information is not collected or shared. This model can serve as a cost-effective alternative to visiting a doctor for preliminary symptom analysis. This can be particularly beneficial for individuals who may be concerned about high healthcare costs.

By providing a more reliable source for symptom analysis compared to general internet searches, the model may help reduce the spread of medical misinformation.

Github link:
https://github.com/smartinternz02/SI-GuidedProject-602573-1697544681/tree/main/Project%20Development%20Phase/Disease-Prediction-ML

7.2 Feature 2

8. PERFORMANCE TESTING

8.1 Performance Metrics

	Training Accuracy	Validation Accuracy	Testing Accuracy
Naive Bayes Classifier	1.0	1.0	1.00000
K Nearest Neighbors Classifier	1.0	1.0	1.00000
Support Vector Machines	1.0	1.0	1.00000
Decision Trees Classifier	1.0	1.0	0.97619
Random Forest Classifier	1.0	1.0	0.97619

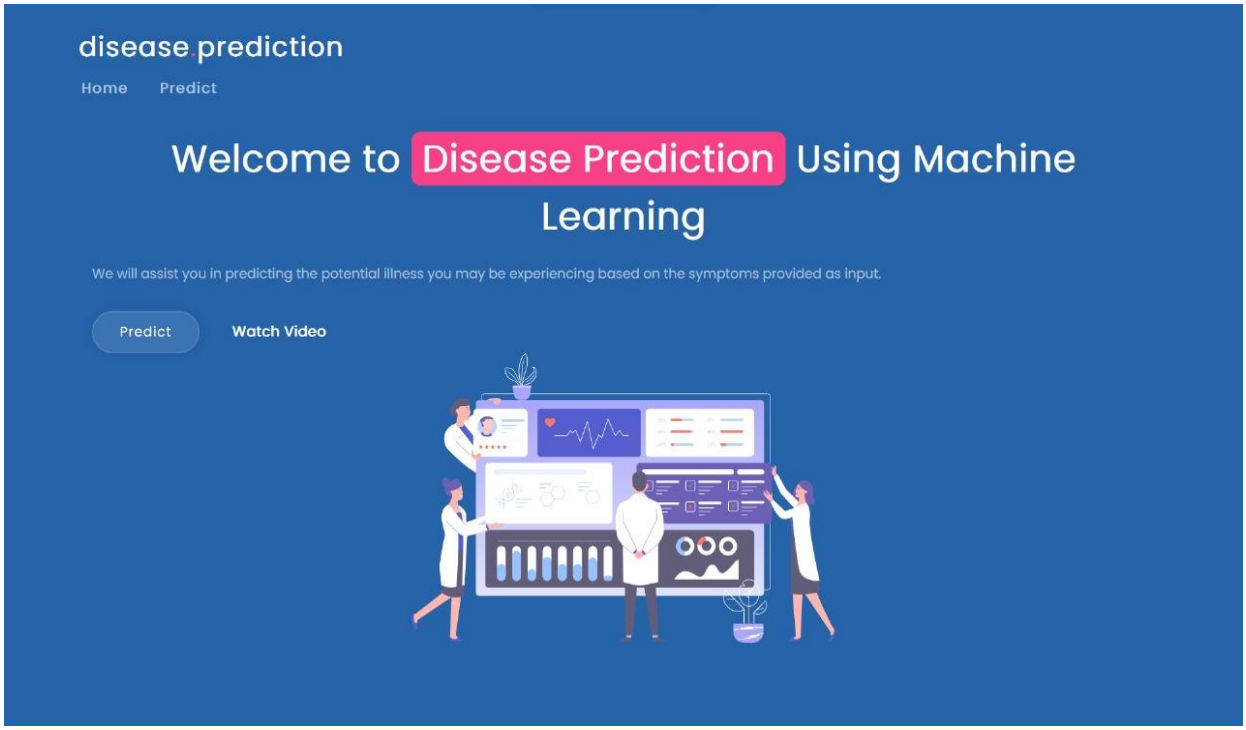
After further tuning

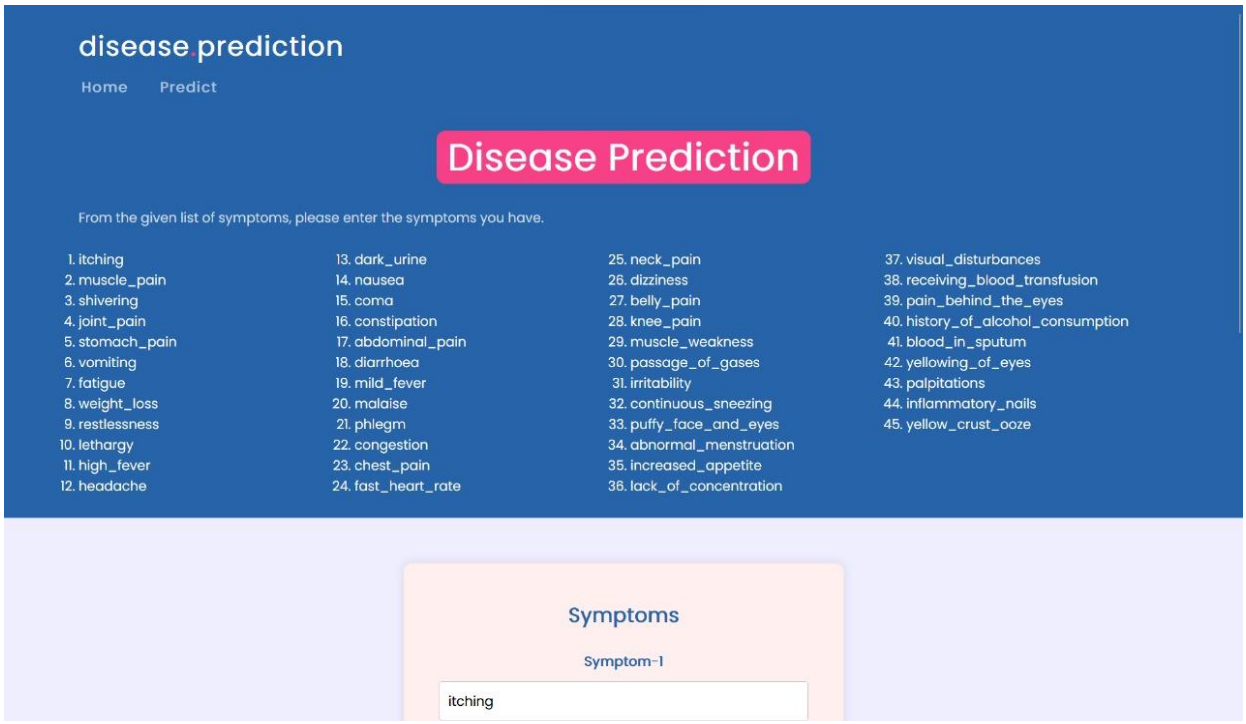
KNeighborsClassifier()

```
The Training Accuracy of the algorithm is  0.8884654471544715
The Validation Accuracy of the algorithm is  0.891260162601626
The Testing Accuracy of the algorithm is 0.9047619047619048
```

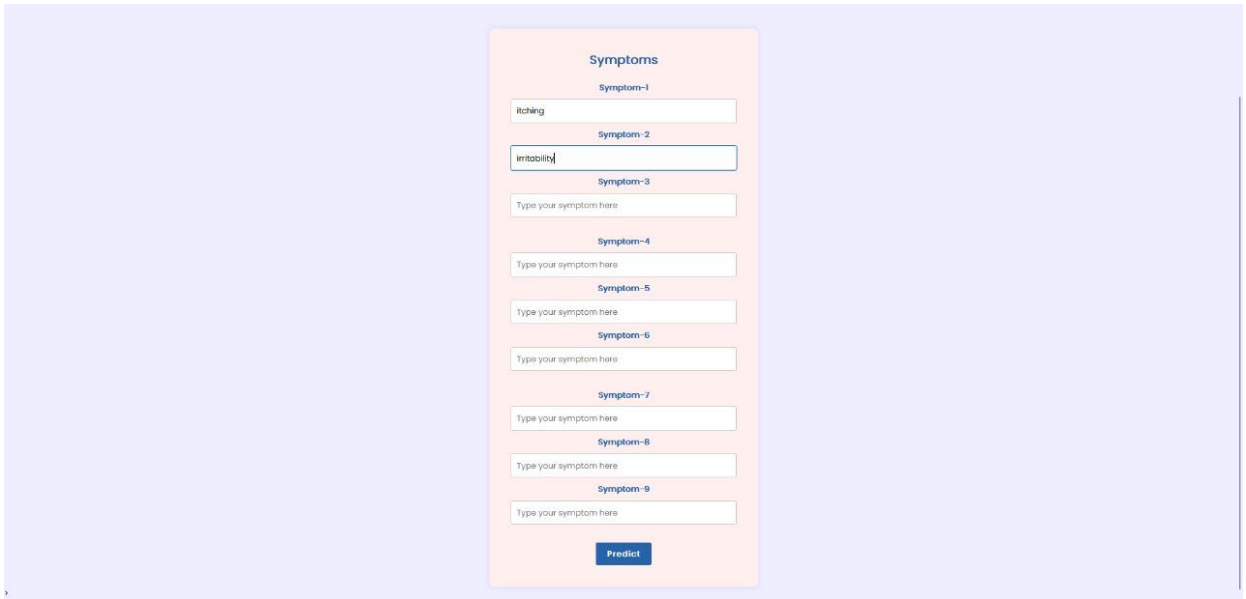
9. RESULTS

9.1 Output Screenshots

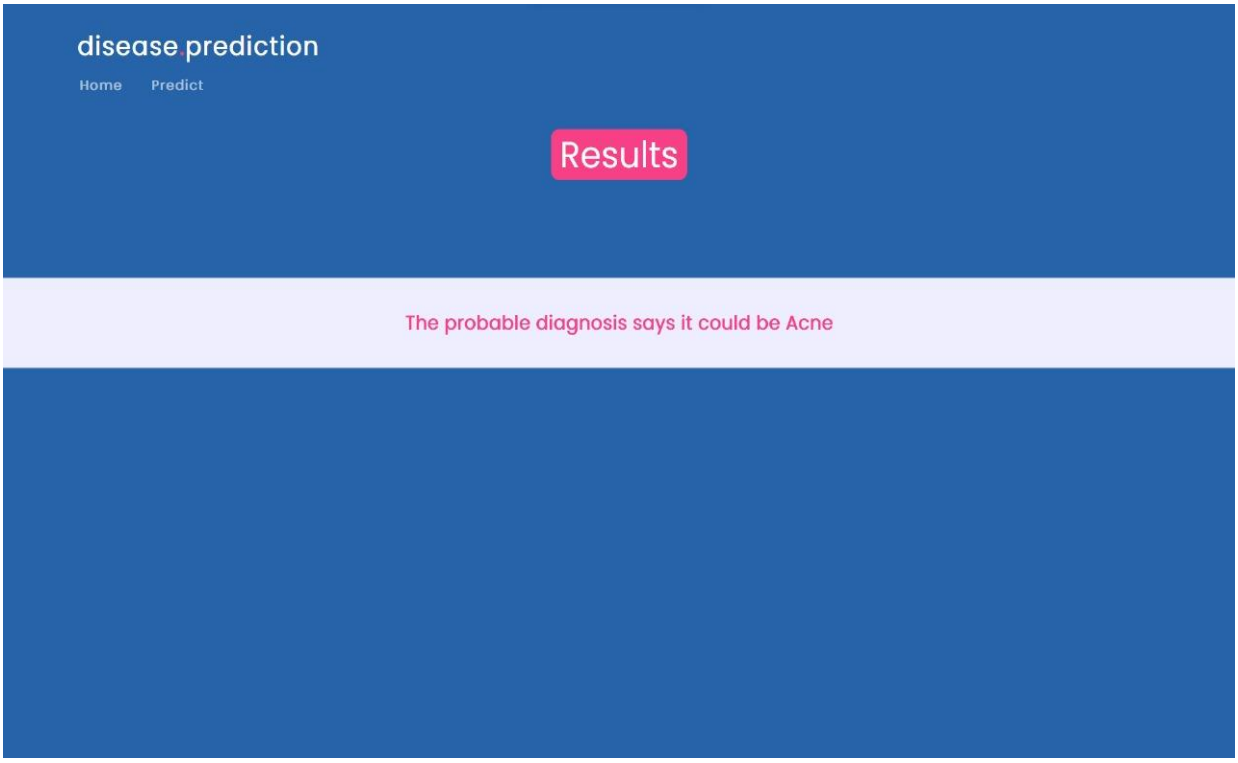




The user/patient can input 9 symptoms. Here, For example, we have added the symptoms of the patients suffering from Acne that are itching and irritability



The result displayed is the patient is suffering from Acne



10. ADVANTAGES & DISADVANTAGES

Advantages:

1. **Efficient Disease Identification:** One of the primary advantages is the efficient and quick identification of diseases. Machine Learning and Deep Learning models can analyze a set of symptoms and provide probable disease predictions, enabling users to take timely action.
2. **Accessibility and Convenience:** This project makes healthcare more accessible to a wider audience. Users can obtain preliminary assessments of their health without the need for immediate in-person medical consultations, which can be time-consuming and costly.
3. **Cost Reduction:** By empowering individuals to perform preliminary health assessments, the project has the potential to reduce healthcare costs associated with unnecessary doctor visits, tests, or consultations. This can be particularly beneficial for individuals without comprehensive health insurance.
4. **Privacy Preservation:** The model operates without the need for users to share personal information, maintaining their privacy. Users can receive predictions based solely on symptoms, without revealing sensitive data.
5. **Use of Rich Data:** Leveraging a dataset with medical images and symptom data from Kaggle allows the model to utilize a diverse and rich source of information, potentially improving the accuracy of disease predictions.

Disadvantages:

1. **Risk of Misdiagnosis:** Machine Learning models, while powerful, are not infallible. There is a risk of misdiagnosis, which can lead to false assurances or unnecessary panic. Users must understand the limitations of the model and consult with medical professionals for conclusive diagnosis.
2. **Dependency on Data Quality:** The model's accuracy heavily relies on the quality and comprehensiveness of the training data. If the dataset contains biases or lacks certain diseases or symptoms, it may lead to inaccuracies in predictions.
3. **No Human Touch:** The model lacks the human touch and expertise of a healthcare professional. While it can provide preliminary assessments, it cannot offer the nuanced, personalized care and guidance that a doctor can.
4. **Limited Scope:** The project's scope is limited to predicting diseases based on a set of symptoms. It cannot address all medical concerns, especially those that require physical examinations or complex diagnostic procedures.
5. **Technical Barriers:** Accessibility to the project may be limited by technical barriers. Users with limited technology access or digital literacy may find it challenging to use the web application effectively.

11. CONCLUSION

In conclusion, the development of a model that can predict up to 42 diseases based on input symptoms represents a significant step forward in addressing some of the challenges faced by individuals in today's fast-paced world. With limited time for healthcare and the often-unreliable search results from internet searches, this model offers a valuable solution.

This model can serve multiple purposes, benefitting both patients and healthcare providers. It allows patients to have a preliminary understanding of potential health issues without the need for an immediate doctor's visit, which can be costly and time-consuming. Moreover, the model respects privacy by not requiring personal data, making it a user-friendly and secure tool.

For healthcare professionals, this model can be an invaluable resource for online consultations, offering a more efficient and informed approach to patient care. It streamlines the diagnostic process, potentially reducing the burden on healthcare systems and ensuring that patients with serious conditions receive the attention they need in a timely manner.

While this model is a promising development, it's essential to note that it should not replace professional medical advice. Instead, it should be seen as a complementary tool, offering initial insights and guidance. The importance of consulting a healthcare provider for a comprehensive evaluation and personalized treatment cannot be understated.

In a world where time and resources are often limited, this model provides a practical solution to address health concerns and make informed decisions about seeking medical care. It represents a step toward more accessible and efficient healthcare, ultimately contributing to better overall health and well-being.

12. FUTURE SCOPE

1. Efficient Doctor Consultation: Doctors can use the model to consult with patients online, potentially saving time for both the medical professionals and patients. This can be especially useful in situations where in-person visits are challenging or unnecessary.
2. Integration with Wearable Devices: With the increasing popularity of wearable health devices, your model could integrate with these devices to provide real-time health monitoring and early symptom detection. This could help in proactive healthcare management.
3. Partnerships with Healthcare Providers: Collaborating with healthcare institutions can enhance the credibility and reach of the application. This can include providing the application as a value-added service to patients.
4. Skin diseases are occurring almost on all groups of ages among people.The rate of skin disease has been increased due to lifestyle and changing environments, Image processing for skin diseases.

13. APPENDIX

Source Code

app.py

```
from flask import Flask, render_template, request
import numpy as np
import pickle


model = pickle.load(open('model.pkl','rb'))
app = Flask(__name__)


@app.route("/")
def home():
    return render_template('index.html')


@app.route('/details')
def pred():
```

```
    return render_template('details.html')


@app.route('/predict',methods=['POST','GET'])
def predict():
    col=['chills', 'joint_pain', 'vomiting', 'fatigue', 'weight_loss',
       'restlessness', 'lethargy', 'cough', 'high_fever', 'sweating',
       'headache', 'dark_urine', 'nausea', 'loss_of_appetite',
       'pain_behind_the_eyes', 'back_pain', 'diarrhoea', 'mild_fever',
       'yellowing_of_eyes', 'blurred_and_distorted_vision', 'phlegm',
       'congestion', 'chest_pain', 'fast_heart_rate', 'puffy_face_and_eyes',
       'excessive_hunger', 'knee_pain', 'muscle_weakness', 'stiff_neck',
       'swelling_joints', 'loss_of_balance', 'unsteadiness',
       'bladder_discomfort', 'passage_of_gases', 'depression', 'irritability',
       'muscle_pain', 'abnormal_menstruation', 'increased_appetite',
       'family_history', 'mucoid_sputum', 'rusty_sputum',
       'lack_of_concentration', 'receiving_blood_transfusion', 'coma',
       'history_of_alcohol_consumption', 'blood_in_sputum', 'palpitations',
       'inflammatory_nails']
    if request.method=='POST':
        inputt = [str(x) for x in request.form.values()]


        b=[0]*49
        for x in range(0,49):
            for y in inputt:
                if(col[x]==y):
                    b[x]=1
        b=np.array(b)
        b=b.reshape(1,49)
        prediction = model.predict(b)
```

```
        prediction = prediction[0]
    return render_template('results.html', prediction_text="The probable diagnosis says it could be
{}".format(prediction))


if __name__ == "__main__":
    app.run()
```

GitHub link:
https://github.com/smartinternz02/SI-GuidedProject-602573-1697544681/tree/main/Project%20Development%20Phase/Disease-Prediction-ML

Project Demo Link:

https://www.youtube.com/watch?v=M6gBsphsxE4