**PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS**

## A PROJECT REPORT

***Submitted by,***

**Kuchi Harika-20211CSE0182**

**Gurramkonda Sandhya-20211CSE0316**

**Siddavatam Navaneeth Karthik-20211CSE0326**

**Pamidimarri Praveen Kumar-20211CSE284**

### *Under the guidance of,*

**Ms. DHANYA D**(Assistant professor ,SoCSE)

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

**BACHELOR OF TECHNOLOGY**

**IN**

**COMPUTER SCIENCE AND ENGINEERING, COMPUTER ENGINEERING, INFORMATION SCIENCE AND ENGINEERING Etc.**

**At**



**PRESIDENCY UNIVERSITY**

**BENGALURU**

**DECEMBER 2024**

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**CERTIFICATE**

This is to certify that the Project report **“PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS”** being submitted by “KUCHI HARIKA” bearing roll number “20211CSE0182” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

|  |  |
| --- | --- |
| **Ms. DHANYA D**  Assistant professor  School of CSE&IS  Presidency University | **Dr. ASIF MOHAMMED**  Professor & HOD  School of CSE&IS  Presidency University |

|  |  |  |
| --- | --- | --- |
| **Dr. L. SHAKKEERA**  Associate Dean  School of CSE  Presidency University | **Dr. MYDHILI NAIR**  Associate Dean  School of CSE  Presidency University | **Dr. SAMEERUDDIN KHAN**  Pro-VC School of Engineering  Dean -School of CSE&IS  Presidency University |

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**CERTIFICATE**

This is to certify that the Project report **“PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS”** being submitted by “GURRAMKONDA SANDHYA” bearing roll number “20211CSE0316” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

|  |  |
| --- | --- |
| **Ms. DHANYA D**  Assistant professor  School of CSE&IS  Presidency University | **Dr. ASIF MOHAMMED**  Professor & HOD  School of CSE&IS  Presidency University |

|  |  |  |
| --- | --- | --- |
| **Dr. L. SHAKKEERA**  Associate Dean  School of CSE  Presidency University | **Dr. MYDHILI NAIR**  Associate Dean  School of CSE  Presidency University | **Dr. SAMEERUDDIN KHAN**  Pro-VC School of Engineering  Dean -School of CSE&IS  Presidency University |

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**CERTIFICATE**

This is to certify that the Project report **“PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS”** being submitted by “SIDDAVATAM NAVANEETH KARTHIK” bearing roll number “20211CSE0326” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

|  |  |
| --- | --- |
| **Ms. DHANYA D**  Assistant professor  School of CSE&IS  Presidency University | **Dr. ASIF MOHAMMED**  Professor & HOD  School of CSE&IS  Presidency University |

|  |  |  |
| --- | --- | --- |
| **Dr. L. SHAKKEERA**  Associate Dean  School of CSE  Presidency University | **Dr. MYDHILI NAIR**  Associate Dean  School of CSE  Presidency University | **Dr. SAMEERUDDIN KHAN**  Pro-VC School of Engineering  Dean -School of CSE&IS  Presidency University |

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**CERTIFICATE**

This is to certify that the Project report **“PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS”** being submitted by “PAMIDIMARRI PRAVEEN KUMAR” bearing roll number “20211CSE0284” in partial fulfillment of the requirement for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a bonafide work carried out under my supervision.

|  |  |
| --- | --- |
| **Ms. DHANYA D**  Assistant professor  School of CSE&IS  Presidency University | **Dr. ASIF MOHAMMED**  Professor & HOD  School of CSE&IS  Presidency University |

|  |  |  |
| --- | --- | --- |
| **Dr. L. SHAKKEERA**  Associate Dean  School of CSE  Presidency University | **Dr. MYDHILI NAIR**  Associate Dean  School of CSE  Presidency University | **Dr. SAMEERUDDIN KHAN**  Pro-VC School of Engineering  Dean -School of CSE&IS  Presidency University |

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled **PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of

**Ms. DHANYA D, Assistant professor,** **School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Student Name Roll number Signature

Kuchi Harika 20211CSE0182

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled **PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of

**Ms. DHANYA D, Assistant professor,** **School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Student Name Roll number Signature

Gurramkonda sandhya 20211CSE0316

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled **PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of

**Ms. DHANYA D, Assistant professor,** **School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Student Name Roll number Signature

Siddavatam Navaneeth Karthik 20211CSE0326

**PRESIDENCY UNIVERSITY**

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**DECLARATION**

We hereby declare that the work, which is being presented in the project report entitled **PREDICTIVE ANALYSIS OF POST-COVID SYMPTOMS** in partial fulfillment for the award of Degree of **Bachelor of Technology** in **Computer Science and Engineering**, is a record of our own investigations carried under the guidance of

**Ms. DHANYA D, Assistant professor,** **School of Computer Science Engineering & Information Science, Presidency University, Bengaluru.**

We have not submitted the matter presented in this report anywhere for the award of any other Degree.

Student Name Roll number Signature

Pamidimarri Praveen Kumar 20211CSE0284

**ABSTRACT**

**The COVID-19 pandemic has left a deep and grave scar and has affected most of Australia's children and their families.**  
Global health is increasingly concerned with the impact of terrorism and climate change on global health in the form of a growing number of climate change scenarios affecting individuals experiencing post-acute sequelae of SARS-CoV-2 in combination with a series of drug-induced epilepsy.

PASC is commonly referred to as post-COVID symptoms such as mild pneumonia (PD), chronic diarrhea (CH), and pneumonia (PC), or long COVID. These symptoms, which range from fatigue and nausea to depression and nausea, mainly occur within a 24-hour period. Cognitive impairments to cardiovascular and respiratory functions have been associated with cognitive impairments and cognitive impairments to cardiovascular and respiratory problems.

Some medical complications pose significant challenges for healthcare systems, patients, and caregivers alike. This research paper delves into the predictive power of prediction techniques. It presents a conceptual study of predictive analytics of post-COVID symptoms using advanced data-driven techniques, including complex genetic and immune testing methodologies, machine learning, and statistical methods.

By leveraging clinical datasets, electronic health records, electronic patient records, and financial data analytics, the study aims to improve patient satisfaction and reduce costs as a whole. It identifies patterns, measures patient-reported outcomes, and examines findings from previous clinical follow-ups. Risk factors and predictive indicators associated with long-term mortality have been identified, as well as various factors related to the risk of injury or death and complications of COVID-19.

Key contributions of this research include the development of a quantitative method for applying findings to real life. Robust predictive models are used to predict the probability and severity of potential biological conditions and their effects. These factors include a clinical history of COVID-19 and its effect on post-infection symptoms, along with demographic and behavioral characteristics. Genetic and clinical variables influencing these outcomes are also examined.

The findings aim to aid in early intervention strategies, personalized treatment plans, and improved resource allocation in healthcare, which are fundamental to better patient outcomes. This paper emphasizes the necessity of integrating fundamental principles of engineering and the development of scientific instruments into the economic and social processes of developing predictive analytics for public health planning to mitigate the long-term burden of the pandemic.

**ACKNOWLEDGEMENT**

First of all, we indebted to the **GOD ALMIGHTY** for giving me an opportunity to excel in our efforts to complete this project on time.

We express our sincere thanks to our respected dean **Dr. Md. Sameeruddin Khan**, Pro-VC, School of Engineering and Dean, School of Computer Science Engineering & Information Science, Presidency University for getting us permission to undergo the project.

We express our heartfelt gratitude to our beloved Associate Deans **Dr. Shakkeera L and Dr. Mydhili Nair,** School of Computer Science Engineering & Information Science, Presidency University, and **Dr. Asif Mohammed**, Head of the Department, School of Computer Science Engineering & Information Science, Presidency University, for rendering timely help in completing this project successfully.

We are greatly indebted to our guide **Ms. Dhanya D, Assistant professor** and Reviewer **Dr./Mr.Ms. Name, Designation**, School of Computer Science Engineering & Information Science, Presidency University for her inspirational guidance, and valuable suggestions and for providing us a chance to express our technical capabilities in every respect for the completion of the project work.

We would like to convey our gratitude and heartfelt thanks to the PIP2001 Capstone Project Coordinators **Dr. Sampath A K, Dr. Abdul Khadar and Mr. Md Zia Ur Rahman,** department Project Coordinators **Dr. Amarnath J L** and Git hub coordinator **Mr. Muthuraj.**

We thank our family and friends for the strong support and inspiration they have provided us in bringing out this project.

**KUCHI HARIKA**

**GURRAMKONDA SANDHYA**

**SIDDAVATAM NAVANEETH KARTHIK**

**PAMIDIMARRI PRAVEEN KUMAR**

**LIST OF TABLES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Table Name** | **Table Caption** | **Page No.** |
| 1 | Table 1.1 | Software modules versus Reusable components | 5 |

**LIST OF FIGURES**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl. No.** | **Figure Name** | **Caption** | **Page No.** |
| 1 | Figure 1.1 | Software modules versus Reusable components | 5 |

**TABLE OF CONTENTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | | **PAGE NO.** |
|  | **ABSTRACT ACKNOWLEDGMENT**  **LIST OF TABLE**  **LIST OF FIGURES** | | **x-xi**  **xii**  **xiii**  **xiv** |
| **1.** | **INTRODUCTION** | | **1** |
|  | 1.1 Basic understanding | | 1 |
|  | 1.2 Problem statement | | 2 |
|  | 1.3 Overview  1.4 Domain Introduction  1.5 Motivation behind the project | | 2  2-3  3 |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
| **2.** | **LITERATURE REVIEW** | | **4** |
|  | 2.1 Introduction | | 4 |
|  | 2.2 Related works | | 4-8 |
|  | 2.3 Existing works | | 8-9 |
| **3.** | **RESEARCH GAP OF EXISTING METHODS** | |  |

3.1 research gap 10

**4. PROPOSED METHODOLOGY**

4.1 proposed method 11-12

4.2Introduction to model &data generation 12-13

4.3 Preparation and data engineering 13-14

4.4 Training strategy 14-15

4.5 Model evaluation and performance 15-16

4.6 Detailed analysis of results 16

4.7 Future development and suggestions 16

4.8 Applications and case studies 16

4.9 Ethical considerations and AI 17

4.10 Challenges overcome 17

**5.**  **OBJECTIVES**

5.1 Synthesis a synthetic dataset 18

5.2 Data preprocessing 18

5.3 Development of ML models 19

5.4 Integration of frontend & backend 19

5.5 Real time symptoms prediction 20

5.6 Evaluation & optimization 20

5.7 Scalability & Ethical implications 20

**6. SYSTEM DESIGN & IMPLEMENTATION**

6.1 React frontend 21

6.2 User flow 22

6.3 API calls 22-23

6.4 Error handling & user feedback 23

6.5 Changing & suggestions 23

6.6 Backend Documentation 24

6.7 How we handle your data 25-26

**7. TIMELINE FOR EXECUTION OF PROJECT**

7.1 Planning and Requirement Gathering 27

7.2 UI/UX Design 27

7.3 System Architecture 28

7.4 Frontend Development 28

7.5 Backend Development 28

7.6 AI/ML Model Development 29

7.7 Testing 29

# 8. OUTCOMES

8.1.Symptom prediction tool: 30

8.2.Data-Driven Insights: 30

8.3.Better Decision Support : 30

8.4.New Applications of Machine Learning 30

8.5.Patient Engagement: 30

8.6.Interdisciplinary Contribution: 30

**9. RESULTS AND DISCUSSIONS**

9.1Results and Conclusion 31-32

9.2Future Directions 32

**10.** **CONCLUSION** ­­ 33

**11. REFERENCES** 34

**Appendix A**

Pseudo code 35-44

**Appendix B**

Screenshots 45

**Appendix C**

Enclosures 46

**CHAPTER-1**

**INTRODUCTION**

* 1. **Basic Understanding**

**The COVID-19 pandemic was caused by the novel coronavirus and caused by the novel COVID.**SARS-CoV-2 has been the defining global health crisis of the past and is now a global public health emergency of the 21st century. While the acute phase of the infection has not yet been detected, the acute phase of the disease has been treated.

With increasing attention being shifted towards more complex studies of biomedical topics and their significance, studies are increasingly concentrated on understanding the long-term health consequences of a poor lifestyle, which is important for doctors. Termed as "Long COVID" or post-acute sequelae of SARS-CoV-2 (PASC), these conditions can include a wide variety of pathogens and are associated with a variety of conditions.

A spectrum of symptoms persists for weeks or months after onset but is also referred to as symptoms that persist for days or months after the initial infection and the resolution of the infection. Fatigue, cognitive dysfunction (cumulatively referred to as Fatigue), and inattention are the major causes of chronic fatigue in brains. Dyspnea and cardiac disorders, the chief causes of brain fog, are the most frequently reported symptoms.

Despite their potential effects, these chronic post-COVID manifestations not only decrease COVID as a social and economic subtype but also make COVID a dread. They affect the quality of life for affected individuals and present a health risk for affected persons. Among other problems, the burden for healthcare systems worldwide has been significant. The health system currently faces an increasing burden for healthcare providers, showing a heterogeneity of symptoms coupled with varying degrees of symptoms.

The corresponding lack of serious evidence suggests the urgent need for a systematic review of existing state intervention plans. There is a need for assessment of individuals at risk for long-range disabilities and their families to determine the risk of long-term disability. These assessments anticipate the trajectory of their recovery, covariates releasing COVID, and predict the trajectory of their recovery.

The economic implications of Long COVID are equally important to consider. They have significant impacts as they lead to reduced productivity among the workforce. A recent analysis found that healthcare costs increased by 3 percent. Further psychosocial aspects are involved. However, the psychosocial aspects are not limited to medical conditions. The impact of long-term illness often exacerbates existing mental health problems.

Health challenges require a comprehensive plan to address their severity and necessitate a multidisciplinary perspective on care for patients.

* 1. **Problem Statement**

the COVID-19 pandemic has resulted in significant long-term health impacts for many individuals collectively termed as "post COVID symptoms" or "long COVID." These symptoms can affect patients' quality of life as well as pose a challenge A preliminary analysis of patients with this rare condition in the early stages could be useful in assessing whether their medical care is effective, and if the symptoms are present in early stages Current methods of identifying post-COVID symptoms often rely on subjective reporting or delayed clinical observations which could not adequately predict long-term outcomes. A data-driven, scalable and automated solution that can predict the probability of specific symptoms based on patient demographics, medical history and severity of COVID-19 infection. This project addresses the problem by developing a predictive web-based API service that leverages machine learning techniques. The system provides based on user input probabilistic predictions for post- COVID symptoms to a trained Random Forest model. This tool is intended to assist healthcare professionals researchers and individuals in understanding potential risks and planning appropriate interventions.

* 1. **Overview**

This project involves the web application which is basically meant for the prediction of the symptoms of post covid or long covid.

It involves a react application, which is a single webpage application which has a form to take a input from user. The form includes a inputs like name, age, gender, fatigue, brain fog, join pain, diabetes, hypertension, breathlessness etc.

It uses flask for the backend which involves machine learning model for prediction. It uses CORS for have cross origin resource sharing.

* 1. **Domain introduction**

This project will fall under a domain healthcare and also under machine learning. Specifically in terms of post-covid symptom prediction. The corona virus or a COVID-19 , the pandemic which has led to increase the focus towards understanding and managing the health for long time from the virus, which we can call as long term covid or a post covid syndrome. These symptoms like fatigue, brain fog and breathless stays even after the infection has passed.  
In the era where the growing of machine learning models such as Random forest are included to predict the probability of various long term covid symptoms based on the factors such as age, gender, hypertension, blood pressure, severity in infection, diabetes etc.

Now, the predictive machine learning models are included with the health care domains for the early predict of the diseases, symptoms etc  
It can also used to determine the health risk of individuals based on demography and the age.  
It can also provide a personalized treatment plans and needing care.  
Our project’s approach for the integration of health care with a predictive model is to provide a predictive tool for the patients and the health care providers. It guides us to understand the risk of post covid symptom and assists in more informed healthcare decisions.

****

Fig 1.1 health care with machine learning domain

* 1. **Motivation behind the project:**

Beyond causing major disruptions to health systems globally, COVID-19 has left the survivors with residual effects. These include the continued symptoms post recovery, including chronic fatigue, difficulty in respiration, and neurologic impairment. These post-COVID conditions can be extremely problematic for individuals as well as their caregivers, so assessment and predictive tools are a prerequisite.This addresses the need by applying the capabilities of machine learning and predictive analytics in order to provide each patient with a tailored assessment of their health.  
The following are the most compelling reasons driving the project forward:  
1. To give early warning in cases of high risk patients towards the development of complications.  
2. To reduce health burdens by allowing patients to take proactive measures beforehand.  
3. To raise general awareness about post COVID condition and management.  
4. Health care with better quality of life using data-driven approach.

**CHAPTER-2**

**LITERATURE SURVEY**

**2.1 Introduction**

Significantly shifting global healthcare priorities, the COVID-19 pandemic has sparked an enormous amount of research on immediate and long-term effects of the disease. Post-COVID problems, commonly referred to as raised COVID, are a field of special interest, where research encompasses various enduring symptoms including exhaustion, mental haze, dyspnea, and joint pain. Such symptoms could drastically reduce the quality of life and might present challenges for healthcare systems across the globe.

A detailed literature review is necessary for understanding the current state of research on long-term COVID and its application in machine learning for symptom prediction. This overview examines studies related to the prevalence, clinical features, and predictive analytics of prolonged COVID. It also describes methods of result prediction and the identification of risk factors through observational research, synthesis studies, and machine learning models.

It examines the ability of data-driven methods like EHRs and health information systems to provide enhanced symptom prediction and classification by exploring their benefits and drawbacks, such as scalability, generalization, and dependency on datasets of high quality.

This paper builds its foundation from a literature review wherein the author shall attempt to make random forest models a part of a predictive symptomatology following the COVID-19 infection. From a basis provided in previous studies' insights, it aims to take a stride ahead in contributing literature to the discipline of predictive medicine and enabling some innovative strategy-making for prolonged Covid.

**2.2 Related work**The rapidly growing prevalence of COVID long COVID in adults has brought a significant increase in the number of cases according to the WHO.

This spurred extensive research efforts to understand its etiology and implications.

manifestations and long-term impacts. Prior studies have examined the impact of a variety of techniques on human health.

Issuing the results shows which are some of the most common cited data points and the most common '' most common ''

symptoms of fatigue comorbidity such as fevers respiratory distress, and dizziness are mainly attributed to lack of motivation to work, when an individual is already sleeping.

neurological impairments. Research by Carfi and colleagues has shown that the majority of the respondents cite carfi and his colleagues as having ''

Over 87% of patients at home were discharged from the hospital within 3 months in 2018 compared to 84% in 2009 the same year.

The most hospitalized patients reported at least one persistent symptom two days a week.

Two months after recovery it was revealed that the patient had developed a thyroid toxicity and was undergoing hormonal imbalances. Similar to Huang et al. (2021) 403-4144 or 632-3100.

The research team conducted a comprehensive cohort study focusing on the association between population age and overall health conditions.

Chronic COVID burden was the biggest burden among patients six years old and younger who had long COVID before the diagnosis of the disease.

A few months after infection the disease was detected and treated.

Machine learning extended COVID-19 symptom analysis

Machine Learning and Artificial Intelligence (AI) have evolved in parallel to the classical AI.

So tools for predictive modeling emerged emerging as powerful tools for research in long-term and short-term evaluation and prediction.

COVID. COVID. For example a study by Davido et al. (2005) demonstrates that a higher-grade RNA is an ideal system for preventing (2021) used by the company (2021) when the employee (2021) was 306

Gradient-boosting algorithms to classify patients based on their age require gradient boosting algorithms to classify patients based on

It also demonstrates significant predictive value at predictive and antidepressant benefit at symptom persistence.

Other important findings from Fang et al.[4] have been that in addition to the proposed model [12] in the first e (2022) used ensemble ensemble ensemble ensemble ensemble ensemble ensemble ensemble ensemble

We use advanced data analysis techniques to analyze clinical and demographic data to better understand human psychiatric health and

The program targets high-risk groups as a result of persistent respiratory symptoms based on clinically relevant clinical information.

These studies demonstrate the potential of AI to address complex problems.

Health challenges associated with Long COVID have been described.

Deep Learning Models such as recurrent neural networks and neural networks can be applied to data in machine learning.

LNs and CNNs have also emerged, convolutional neural networks have been proposed or have been described. These include NNNs and CNNs.

A methodology was applied to predict the onset and severity of post-covarid postimmuno mortality post

symptoms. Work by Chen et al. (2022) highlighted the fact that this day has been ignored in the course of the debate.

Use of temporal data collected from electronic data collection is found to be more effective than using temporal data from analog data

With neural networks to achieve high accuracy of health data and high reliability from all sources at high quality for clinical data has been demonstrated. These networks are used for the

Similarly, transformer-based predictions are used for symptom prediction [8] and transformer-based prediction has been described in [12]

Architectures have been explored for capture of temporal images and spatial features.

The recovery patternsinpatient recovery trajectories have been reported by Zhou et al.

al. (2023).

Statistical methods remain integral to predictive analysis and have been used for many decades.

Cox proportional hazards models and multivariate logistic regression models and Cox proportional hazards models were developed and tested in the Cox

How is regression frequently employed to assess risk factors for Long-Term Death syndromes or depressions?

COVID. COVID.Si. For instance studies by Peluso et al.[5] (2021) have shown that a study of human osteoarthritis has been described[

This study demonstrated the utility of these models in linking in terms of integration of new models to existing model work.

chronic disease such as diabetes and obesity with prolonged periods of high blood pressure and high blood pressure can lead to significant complications such as diabetes and

recovery times. Bayesian frameworks have also been employed in the Bayesian frameworks.

To integrate uncertainty into models of predictions the modeling process provides uncertainty coverage.

Data will help clinicians assess confidence intervals for decision-making.

Comprehensive reviews such as those of Aiyegbusi et al. have been found to be a useful tool for preventing blind

(2021), have shown the multifactorial nature of Long Term Long Term Research.

COVID and the need for interdisciplinary approaches. This course and these books will be of interest to you.

Reviews emphasize the role of psychosocial factors including gender and sexual orientation in the cognitive development of people.

Mental health status and socioeconomic conditions are a major concern to many people in India. These are two aspects of chronic illness that affect the population's mental health

In predictive models predictive models are often overlooked underexplored if the prediction is not already completed. Incorporating such a component in such incorporation requires inclusion of such a device as a personal computer in a process of

Add dimension into ML and statistical models can enhance their precision in ML and statistical models by using statistical variables and a combination

Robustness and applicability are key aspects for a robustness and applicability in real-world settings.

Although progress has been made in the field several challenges remain in the region.

Fields. Many studies are constrained by heterogeneity of studies in the literature.

Data datasets are small sample sizes and inconsistent definitions of the form of data are the main reasons for the large sample size.

Long COVID is an invasive COVID. Moreover the limited availability of refurbished products in the market is a disadvantage in this regard.

Longitudinal data hampers efforts to understand the full potential of longitudinal data.

Addressing these problems will help to solve the problem of symptom progression[6].

The identification of limitations requires that collaborative efforts be made to establish the limits.

The new protocols will be standardized and the new initiatives will expand shared data.

This paper builds upon existing research to improve the characterization of complex environments and new data.

The research will be used as a hybrid approach combining ML database and datasets employing data mining and hybrid approaches.

Statistics, psychosocial factors and statistical techniques [13]. The aim of the project is to support the project and its implementation.

This task provides a comprehensive framework for predicting and evaluating change in human population.

The management of long COVID and COVID managing Long COVID impacted the management of COVID, ultimately contributing to improved COVID control

Patient outcomes and healthcare strategies are critical towards improving patient outcomes and healthcare strategies.

**2.3 Existing works**

| **Sl. No.** | **Paper Title** | **Method Used** | **Advantages** | **Limitations** |
| --- | --- | --- | --- | --- |
| 1 | WHO Coronavirus (COVID-19) Dashboard [(accessed on 25 February 2023)] | Global data collection and visualization platform | Provides real-time data on COVID-19 cases, deaths, and vaccinations worldwide | Limited to data reported by countries, potential inconsistencies in data quality |
| 2 | Global Prevalence of Post-Coronavirus Disease 2019 (COVID-19) Condition or Long COVID: A Meta-Analysis and Systematic Review | Meta-analysis and systematic review of global long COVID studies | Comprehensive global estimate of long COVID prevalence | Variability in study definitions of long COVID, limited generalizability due to study heterogeneity |
| 3 | Prevalence and clinical features of long COVID from omicron infection in children and adults | Observational study analyzing prevalence and clinical characteristics | Focuses on both children and adults, provides insight into Omicron-related long COVID | Limited data for non-Omicron variants, relatively recent analysis with limited long-term follow-up |
| 4 | Identifying who has long COVID in the USA: A machine learning approach using N3C data | Machine learning (ML) models trained on large-scale electronic health records (EHRs) | Innovative use of ML for identifying long COVID, scalable and adaptable for other conditions | Dependence on quality and completeness of EHR data, potential biases in ML models |
| 5 | Basic characteristics and representativeness of the German Disease Analyzer database | Database-based observational study | Well-characterized database for clinical and pharmacoepidemiological studies | Limited to German population data, may not represent other countries or regions |
| 6 | Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme, 10. Revision, German Modification, Version 2023 | Classification system (ICD-10-GM) | Standardized coding for diseases, ensuring consistency in data collection | Requires constant updates to remain relevant with emerging diseases |
| 7 | Impfdashboard Deutschland | Time-series analysis of vaccination data | Provides detailed, up-to-date vaccination statistics for Germany | Limited to Germany, dependent on national reporting accuracy |
| 8 | Anzahl und Anteile von VOC und VOI in Deutschland | Surveillance and genomic analysis | Tracks variants of concern (VOC) and variants of interest (VOI) in Germany | Specific to Germany, limited insights into global trends |

**CHAPTER-3**

**RESEARCH GAPS OF EXISTING METHODS**

The existing literature in post-COVID conditions and predictive healthcare has substantial advancements in the understanding and management of long COVID. However, there are key research gaps to be addressed in further improving the prediction accuracy and overall healthcare outcomes.  
  
Global prevalence studies are useful in understanding the general patterns of long COVID symptoms, but they often suffer from heterogeneity in data collection methods and inconsistent definitions of long COVID. This variability limits the development of universally applicable predictive models. Observational studies have identified correlations between demographic and clinical factors with long COVID outcomes, but they lack the capacity to provide individualized or symptom-specific predictions.  
  
The models of machine learning are promising advances in the identification of long COVID patterns from large datasets but face limitations with respect to the completeness, quality, and biases of the data. For example, less representation of some populations or regions with limited healthcare facility may lead to unfair results. Most existing models are focused on the identification of long COVID cases rather than the prediction of the likelihood of specific symptoms, leaving a huge gap in symptom-targeted predictive tools.  
  
In addition, a well-standardized coding system still exists from the existing disease classification systems, yet in many cases, it is too coarse to class the diverse and complex symptoms of long COVID. This hinders further efforts to create more detailed and symptom-specific predictive frameworks. Filling these gaps will be necessary for building robust, personalized tools for predicting and managing long COVID symptoms among diverse and underserved populations.

**CHAPTER-4**

**PROPOSED METHODOLOGY**

The proposed system will works on developing a predictive framework that finds and helps in managing postCOVID symptoms. It will use the algorithms in machine learning, advanced statistical modeling techniques, and a complete set of data to help deepen the understanding and treatment of Long COVID.7 The key elements of the proposed system are as follows:

**4.1 Proposed method**  
**4.1.1. Data Collection and Integration:**

Aggregating data from various sources, including EHRs, patientreported outcomes, demographic information, and genetic profiles. Data sets with longitudinal designs that track how the disease advances over time.

**4.1.2. Preprocessing/Feature Engineering Standardization:**  
of and cleaning datasets; removing all noise and getting them to appear the same. Extracting the features like age, gender, hypertension, fatigue, and socioeconomic factors. Use of machine learning models for building random forest with scikit or tensorflow, boosting machines, and support vector machines, to predict the onset and severity of post-COVID symptoms.

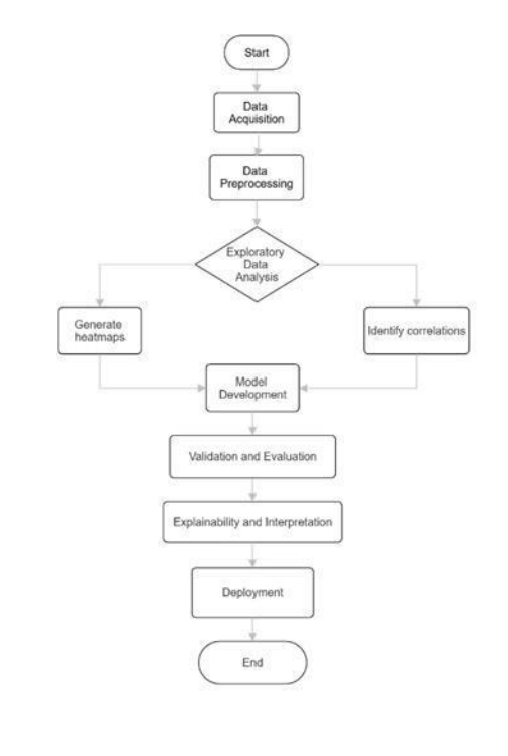
**4.1.3.Predictive model development:**

Using deep learning architectures like LSTM networks and transformer models to capture temporal and sequential patterns in longitudinal data.Future: These can be used in combination for the next advancement with the analysis of multiple machine learning predictions.  
**4.1.4. Verification and Optimisation:**

Using cross-validation methodologies, validate the performance of the model developed to demonstrate its usability. Precision, recall, F1 score, and area under the receiver operating characteristic curve are measures of accuracy of predictive ability of a model. Fine tuning of parameters with grid search and Bayesian optimization.

**4.1.5. Interpretability and Visualization:**

Using explainable AI, SHAP (SHapley Additive exPlanations), with the prediction. Provide simple, understandable dashboards and visualizations to present predictions and risk assessments to the clinicians and patients.  
**4.1.6. Implementation and Deployment: Deployment of the system:**  
Cloud-based platform available to healthcare providers and researchers  
Compliance with data privacy and security standards, including GDPR and HIPAA.



**4.2 Introduction to the Model and Data Generation :**

**4.2.1. Purpose of the Model**

This model will predict the probability of some symptoms that occur after COVID, like fatigue, breathlessness, and brain fog, which are provided with a set of input features like age, gender, diabetes, and hypertension. These are common symptoms that can even persist after recovering from COVID-19, and early detection and monitoring will be important for better management of healthcare.

**4.2.2. Synthetic Data Generation**

In many real-world applications, particularly in the healthcare domain, gathering vast amounts of real data is very challenging mainly due to invasion of privacy, lack of accessibility, or the cost that goes into gathering such data. The synthetic data generated, therefore, emulates the real-world data but does not require real patient information.

The above scenario creates a data generation process with the mimic of the key drivers of post COVID symptoms, namely:

• Age: Age as a factor governs the risk of experiencing any symptom. Older age groups will be more exposed to symptoms of fatigue, breathlessness, and sometimes brain fog

• Gender: Health outcomes tend to vary differently with gender-based differences, specially when an illness subsides. Here, gender forms a categorical variable that is classified as Male and Female.

•Diabetes and Hypertension: They are the most common comorbidities that put a person at higher risk for post-COVID symptoms. For example, an individual suffering from diabetes or hypertension is likely to have more severe or longer-lasting post-COVID symptoms.

These are the fundamental factors that cause the manifestation of symptoms like fatigue, breathlessness and brain fog:

•Fatigue: more commonly occurring in elderly and diabetes.

•Breathlessness: most often seen in elderly or hypertension.

•Brain Fog: A symptom which may persist at even after recovery from COVID-19 especially in elderly patients.

These interdependencies create the artificial dataset such that each symptom produced is dependent on its respective factors.

**4.2.3 Feature Distribution**

All features are simulated using normal distribution or by random choices, clipping, and conditional probabilities are added to create realistic interdependencies in the real world:

•Age is generated from a standard normal distribution with mean 45 years and standard deviation 15 years, clipped to be between 18 and 85 years.

•Gender is assigned randomly with equal chance for Male and Female.

• Diabetes and Hypertension are created as binary (0 or 1) with 20% and 30% respectively.

**4.3 Preprocessing and Feature Engineering**

**4.3.1 Data Cleaning and Preprocessing**

Once synthetic data is created, cleaning and preprocessing happens to prepare it for model training:

•Encoding: Gender is a categorical variable-Male/Female-and must be encoded into a numerical representation to be acceptable for processing by the machine learning algorithm. Here, one-hot encoding is used. This introduces another binary feature for the categorical variable (Male).

•Feature Scaling: Generally, the performance of machine learning models is much better if input features are within a similar scale. All of the above features, age, diabetes, and hypertension are standardized by StandardScaler. During the transformation, features will automatically get aligned to have a mean of 0 and a standard deviation of 1 that makes training efficient and avoids any feature from dominating because of scale.

**4.3.2 Feature Selection and Engineering**

Feature engineering is the application of domain knowledge for creating new features or transformation of existing ones with the intention of improving the performance of a model. This dataset contains a reasonable number of features, but all of them are not used for predicting the target variables-fatigue, breathlessness, and brain fog. Thus, those features that happen to be redundant-age, diabetes, hypertension in the case at hand-are removed from the feature set before training the model.

More interaction features can be incorporated. For example,

Interaction feature between age and diabetes. This could represent the idea that patients of older ages who have diabetes are more responsive to some symptoms.

To make things easier, however, the feature set here is very basic and only includes the following:

Age

Gender

Diabetes

Hypertension

The output, or target labels, of the studied model are the presence or absence of symptoms like fatigue, short breath, and dizziness. Such problems are of a class type that is binary.

**4.4. Training Strategy**

**4.4.1. Model Type**

For this task, Random Forest Classifier is used. Random Forest is an excellent ensemble learning technique that creates a forest of several decision trees and then averages out their predictions to have better accuracy against overfitting.

•Why Random Forest?

Handling Non-linearity: Because the relationship of symptoms to the factors involving age, diabetes, and hypertension is not precisely linear, a random forest might be able to handle the non-linear relationship between features.

Multioutput: It's possible for the classifier to respond to more than one target variable, which puts this in the case of the multioutput classification since three different outputs have been detected-fatigue, breathlessness, and brain fog.

Overfitting: generally, the number of overfitting tends to drop on artificial data by using the random forests when there are averages of multiple decision trees

**4.4.2 Model Training**

• Training Process: Use X\_train (preprocessed features) in training along with the target y\_train to train the model by fitting multiple decision trees to subsets of training data and combining output from all the fitted trees into final output.

•Hyperparameters: Here the model is initiated by n\_estimators=100 i.e. 100 trees are built in the forest. To further tune and improve the model's performance more hyperparameters including the maximum depth the tree must have and minimum sample count it must have before further splitting the node, etc. can be tuned

**4.5. Model Evaluation Metrics and Performance**  
**4.5.1 Evaluation Metrics**  
There are following evaluation metrics used for calculating the performance of this model:  
• Accuracy: The ratio of the correct predictions of positive observations with the total predictions made positive. It answers to the question "Out of how many correctly predicted positives, are actually positive?"  
• Recall: The ratio of correct prediction of positive observations towards the actual number of positives in a case. It answers "How many out of the actually positive ones are correctly identified?"  
• F1-Score: It is a harmonic mean that is balanced for precision and recall.  
• Actual: count of how often a class occurs in a data.

In the classification report, these parameters are mentioned for all the three complaints are as follows.  
+fatigue  
symptom: shortness of breath,  
symptom : confusion.

**4.5.2 Model Outcomes**  
All the model output classifications are tested against the original labels (y\_test).  
Carries out the classifications for each symptom:  
•Fatigue: Compare how well the model is regarding the amount of that exists with the amount that fatigue is there of.  
•Breathlessness: Predictive ability in breathlessness on the model side is tested also  
•Brain Fog: At last, performance concerning the brain fog prediction from the model's perspective is checked  
**4.6 Detail Analysis on Result**  
In comparison to classify reports, and identify where exactly does the model well and requires the improvement as follows  
•Fatigue: If the model performs well relative to the case of fatigue prediction, then this feature set-a combination of ageing and diabetes is significantly affecting fatigues are well represented with the model  
•Breathlessness: If this model is unable to perform too well with regards to breathlessness; it could say that there is additional feature set/representation needed relative to hypertension.  
• Brain Fog: If the forecast of such brain fog is not very accurate, then more feature engineering and deep understanding of such determinants will be required.

**4.7. Future Development and Suggestions**

**4.7.1 Hyperparameter Optimisation**

•Grid search or randomised search in the process of hyperparameter tuning with regards to optimum parameters for the identification of a most apt random forest model.

•Class Imbalanced: If such imbalanced sets of classes occur within your data, either less number of cases present with acute manifestations, or, in here again, comes in class weighting/SMOTE. This acronym defines Synthetic Minority Over-sampling Technique.

**4.7.2 Alternative to Advanced Model**

•Gradient Boosting: In such a scenario, XGBoost or LightGBM might be more appropriate, especially in scenarios of imbalance or when there are non-linear interplay between the variables.

•Neural Networks: If relationships are complex, this may be attempted using a deep learning approach

**4.8 Applications and Case Studies**

This model can be used to identify who is likely to have symptoms after COVID. If so, the diagnosis would be much earlier, thus better monitored and treated, possibly affecting the long-term predictions more favorably.

**4.9. Ethical Considerations and AI Used Responsibly**

Medical data-even calls for even more vigilance in terms of ethical considerations:

•Data Privacy: All data of reality which is utilised during training the model have to be anonymised and secure

•Bias: Models may create bias. Like if synthetic data cannot represent all kinds of population, predictions by model turn out to biased.

**4.10 Challenges Overcome:**

The Post-COVID Symptom Risk Predictor was developed after overcoming the following challenges:

**4.10.1.Availability of Data:**

The data related to post-COVID conditions is sparse and sometimes not very reliable. In order to handle this problem, the project employed synthetic datasets. These were constructed in a manner that would resemble real-world distributions and correlations.

**4.10.2.Model Precision:**

Precision in prediction was an essential aspect. The Random Forest model was trained and tested repeatedly for it to achieve consistent performance.

**4.10.3. End User- friendliness:**

It had to be created in an environment with diverse tech expertise. Simple, user friendly interface has to be created about the React front-end.

**4.10.4. Scalability:**

The program has to make provision for upcoming changes like expansion of symptoms plus the finer aspects of the predictors.

**CHAPTER-5**

**OBJECTIVES**

**5.1 Synthesize a Synthetic Dataset**

This phase aims at the creation of synthetic data to resemble real patterns for post-COVID symptoms. A dataset acts as the training source for the predictive machine learning model on the grounds of fatigue, breathlessness, and brain fog, so reliable analysis and effective predictions are achievable.

Features of the dataset:

Age. The numeric feature displaying mixed range between 18 and 85 years of population

Gender: a categorical feature with male and female.

Comorbidities: A set of two binary features- diabetic and hypertension.

Symptoms: two binary values associated with covid related symptoms such as tiredness, breathlessness and brain fog

Long COVID: This is one expected feature generated due to existence of more than one symptom above

Key Deliverables:

Good and accurate reliable dataset

attributes which include Gender, Age, Hypertension etc.

**5.2 Data Preprocessing**

The data is appropriately preprocessed and cleaned for use in training the model, which entails detailed preprocessing as below:

Noise Removal: Deletion of erroneous or irrelevant entries

Handling Missing Values: Filling missing data or rejection of incomplete records

Feature Scaling: Scaling of numerical attributes to have uniformity

Categorical Encoding: Categorical variables such as income brackets encoded into machine-readable formats

Implementation Steps

➢ numpy and pandas used for preprocessing

➢ Validation of the preprocessed dataset with statistical measures and visualizations.

**5.3 Development of Machine Learning Model**

The system employs Random Forest Classifier machine learning model for the detection of long term covid symptoms percentage. The following are the steps:

➢ Data Splitting: Split the dataset into the training and testing sets using train\_test\_split.

➢ Model Training: Train model using Random Forest Classifier.

➢ Model Evaluation: Assess the model in terms of accuracy, precision, recall, and F1-score.

Implementation Tools:

➢ Libraries used : numpy, pandas, scikit-learn.

➢ Algorithms: Random Forest for good accuracy and scalability.

**5.4 Integration of Frontend and Backend**

The system is composed of a React frontend and Flask backend which will make it easy for the users to interact and detect the percentage of long term symptoms:

Frontend (React)

➢ Easy input forms for input submission to the random forest model.

➢ Live results with the percentage of a few symptoms.

Backend (Flask)

➢ REST APIs to take user inputs and process data.

➢ Using joblib loads the pre-trained random forest model.

➢ Using api calls between the frontend and backend for resource sharing.

Result:

➢ Working web application for long term symptom prediction.

**5.5 Real-Time symptoms prediction**

The system will give fast and accurate predictions.

➢ Low Latency: optimize data transfer between frontend and backend.

➢ Scalability: prediction is accurate for a different quantity of user input.

**5.6 Evaluation and Optimization**

Ensure the system tested and optimized: it is not prone to being unreliable

Cross-Validation Evaluate the model: check the system's performance

Hyperparameter Optimize Random Forest parameters

Hyperparameter Tuning System Test - test the app under various settings

System testing

Outcome:

A sound, efficient as well as highly accurate fraud-detection system.

**5.7 Scalability and Ethical Implications**

The system should be able to scale up for increasing data volume and accommodate future improvements:

➢ Scalability: It must be able to process large amounts of data and be able to accommodate more machine learning models.

➢ Ethics: Synthetic data must not lead to biased or unethical profiling.

**CHAPTER-6**

**SYSTEM DESIGN & IMPLEMENTATION**

**6.1ReactFrontend**  
The "Post-COVID Symptom Risk Predictor" front end is built with React, which is a very efficient and flexible JavaScript library in building interactive user interfaces. It serves as an app that forms the user interface to access the predictive model; this way, frictionless interaction between users and predictive models happens while putting emphasis on both

user experience and accessibility.  
This paper has a deeper structure, design, and technical choice of implementing React frontend with specific importance toward an effective and more usable and effective system.  
Core Structure:  
A central piece inside the App.js file represents that of interaction at the highest component level by allowing the states present in those several components or application programming

interface at the highest extensile manner. Thus it is mainly characterized by:

**6.1.1. User Input Capture:**  
o Form fields for age, gender, and symptom history of the user.  
o Drop-downs and radio buttons for categorical input-diabetes, hypertension, and COVID severity.  
**6.1.2. API Interaction:**  
o Axios based POST requests to Flask backend for the predictions.  
**6.1.3. Result Presentation:**  
o Dynamically the result of prediction with particular user feedback.  
**6.1.4. Error Handling**  
o Mechanism to handle API calls error-incorrect input from the user side or server side is unavailable.

**6.1.5.Styling:**  
It has used both Tailwind CSS and custom styles for beautiful and responsive design. It sets clarity and user-friendliness first so that users can easily find their way around the application.  
•Consistent: Color scheme, typography, and spacings have been utilized uniformly.  
•Responsive: It supports multiple screen size resolutions to be convenient and accessible on any device.

**6.2User Flow**  
The user flow of the application has been optimized to have low cognitive load.  
1. Input of their name and demographic details  
2. It starts filling in health-related questionnaires using extremely intuitive form inputs.  
3. When submitted, it shows the outputs of the predictions with some actionables, though disclaimers do prevail.  
User Inputs, State Updates, and APIs  
This application uses controlled components in React for handling user input and ensuring that with all the form elements, the state is updated accordingly. For every input field, the value and an onChange handler are assigned so that state updates can reflect in real-time.  
Example  
<input  
type="number"  
name="age"  
value={formData.age}  
onChange={handleChange}  
placeholder="Enter your age"  
required  
className="w-full p-2 border rounded"  
/>  
This application makes use of the useState hook of React for handling state. Some of the major state variables are:  
formData : It holds the user's input.  
name: It takes the name of the user.  
prediction: It keeps the response provided from the backend.  
loading: It keeps the status of the API request.  
error: It keeps the error, if anything happens while calling the API.  
**6.3API Calls:**  
Axios is used for POST requests to the Flask backend in the application. The handleSubmit function processes the form data to call the API.  
const handleSubmit = async (e) => {  
e.preventDefault();  
setLoading(true);  
setError(\"\");  
setPrediction(null);  
  
try {  
await axios.post("http://localhost:5000/predict", formData);  
setPrediction(response.data);  
} catch (err) {  
setError(err.response?.data?.error || "Prediction failed");  
} finally {  
setLoading(false);  
};  
This code will ensure that an error handling and feedback mechanism are strong, thereby alerting the users with clear messages in case of a failure.  
**6.4Error Handling and User Feedback Mechanisms**  
**6.4.1Error Handling:**  
The application anticipates errors such as server error or invalid input and has a complete mechanism of error handling:  
1. Validation: It validates all the required fields before the submission.  
2. API Response Handling: It handles backend response errors and shows appropriate messages.  
**6.4.2User Feedback:**  
The application provides user feedback to enhance user engagement and confidence:  
• The loading indicators indicate to the user that their request is being processed.  
• Success messages display predictions in a very simple form.  
• Error messages include practical actions such as checking the inputs again, or trying again some other time.  
**6.5 Changes and Suggestions**  
**6.5.1.Validation**  
o More validation on the input fields. This could also include age range check or even disallow nonnumeric input.  
o Show feedback about the validation to the user in real time.  
**6.5.2.Accessibility**  
o Be friendly to the users who will be using assistive devices by using ARIA roles and labels.  
ensure the page is navigated by both keyboard and screen reader for friendliness.  
**6.5.3.Style Enhancements**  
Further enhance the visuality and then make error or predictions nearly even more efficient.  
 Dark mode support for better usability in low-light conditions.  
**6.5.4.API Integration**  
works only on API call failure  
informative error logs while developing the solution and debugging the same  
**6.5.5.User Engagement:**  
o Progress bar to help the user through the inputting process.  
o Tooltip/help icons to show complicated terms and options.  
**6.5.6.Scalability:**  
o Modular design for components so they can be easily reused and maintained.  
o Use state management libraries for complicated scenarios; Redux or Zustand are some examples.  
**6.5.7. Localization and Multilingual Support:**  
o Support multiple languages to cater to a diverse user base.

**6.6Backend Documentation**

**6.6.1Introduction**

Our backend is the core of the Post-COVID Symptom Risk Predictor, built with Flask. We have managed to make it user-friendly with robust functionality - from data processing to delivering predictions that help understand potential post-COVID symptoms.

API Endpoints

We kept our API structure simple and intuitive with two main endpoints:

1. Health Check (`/`)

Want to check everything is up and running? Just make a GET request to our root endpoint. You will receive a response in this format:

json

{

"service": "Post-COVID Symptom Predictor",

"version": "1.0.0",

"status": "operational"

}

**6.6.2. Prediction Endpoint (`/predict`)**

This is where the magic happens. Send us your patient data with a POST request and we'll return their risk predictions in plain English:

json

{

"predictions":

"fatigue": "80% likelihood of long-term fatigue.",

"breathlessness": "50% likelihood of breathlessness.",

"brain\_fog": "30% likelihood of brain fog."

},

"disclaimer": "This is a predictive tool and not a medical diagnosis. Always consult healthcare professionals."

}

**6.7How We Handle Your Data**

**6.7.1 Input Validation**

We are serious about data quality. We check the following before making any predictions:

- All relevant information is included (age, diabetic status, gender, etc.)

- The data types match what we would expect (numbers for age, yes/no for conditions)

- The values make sense (for example, age between 18-85 years)

**6.7.2 Data Preprocessing**

After validation, we get our data ready for our model by following these steps:

1. Translate your input to the format that our model expects

2. Translate categorical data, such as gender, into a format our model can use

3. Rescale numeric values to make predictions comparable

**6.7.3 The Prediction Process**

What Works Behind the Scenes

We have these three most important parts to produce our predictions:

Trained Random Forest model ( random\_forest\_model.pkl )

Data scaling tool ( feature\_scaler.pkl )

feature reference guide ( feature\_names.pkl )

How We Make Our Predictions

1. Clean input data; prepare it for safe processing

2. Our Random Forest model computes risk probabilities

3. We take that probabilistic output and convert to a simple, actionable message

For instance, rather than exposing users to numbers and figures, we will indicate whether there's a "long term chance of 80% from experiencing long fatigue."

Scalable Application

Scalable Capabilities

We designed and prepared for increased use on:

Setup Deployment across multi-severs easy configuration

Deployment consistency with the application using Docker

Task queue capabilities to absorb hectic periods in terms of usage. It optimizes queries done on the databases.

Optimized Performance

How we manage high performance; to keep our activities running. On data process,

We minimized long processes about Data

Cachings Predicted common response.

- Optimized our model to better use memory space

- Watch performance for catches and bottlenecks so they can be remedied before being a problem

Future

Our system is good now, but not perfect. Next updates will consist of:

- Real-time process improvements

- Advanced caching

- Model optimisations

This will enable us to handle many more users yet still produce our predictions fast and accurately.

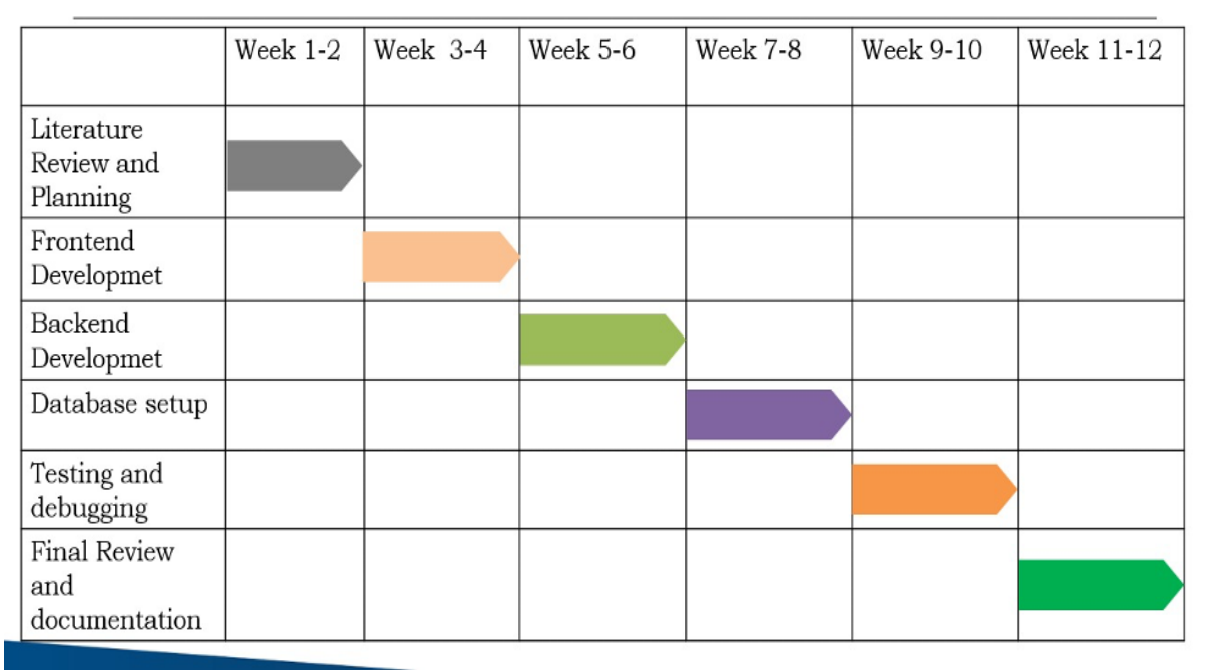
Last Thoughts

Our backend is a combination of medical expertise and technical innovation that will give you reliable post-COVID symptom predictions. While we're proud of its accuracy, remember that it's designed to support, not replace, professional medical advice.

**CHAPTER-7**

**TIMELINE FOR EXECUTION OF PROJECT**

**(GANTT CHART)**

****

**7.1 Planning and Requirement Gathering (Sep 9 – Sep 15, 2024)**  
Define post-covid symptom prediction system objectives and requirement.  
Discussion with Stakeholders: Establish needs from health care professionals, patients and others who will make use of the symptom prediction well.  
Documentation: Highly detailed input specifications such as age, gender, and symptoms that are to be matched against by the desired output such as a probability of occurrence of the symptoms.  
Planning Tools: Gantt charts and project management software; create and manage task dependencies.  
  
**7.2 UI/UX Design (Sep 16 – Sep 22, 2024)**  
Design: Intuitive interface to input symptom data and visualize the results.  
Paper prototypes: Make interactive wireframes to design an intuitive and visually appealing interface with Figma or Adobe XD  
Usability testing: Recruit feedback on the prototypes to ensure that it is user-friendly  
Outcome: Completed UI/UX design on React to use for development.

**7.3 System Architecture (Sep 23 – Sep 29, 2024)**

Describe the general architecture of the project  
Architectural Patterns: Client-server architecture. Flask will take over backend and React as frontend  
Database Schema: Features and storage requirements in the database to scale  
Integration Plan: How frontend React, Backend Flask and the machine learning model would interface with each other  
Outcome: Architecture that is robust, data flow clear and module dependencies well defined.

**7.4 Frontend Development (Sep 30 – Oct 13, 2024)**  
Develop the User Interface using React.  
Component-Based Design: Create reusable React components for input forms, result visualization, and loading states.  
Style Framework: Apply CSS frameworks such as Bootstrap or Tailwind CSS for responsive design  
Cross Browser Compatible: Test interface on many browsers to ensure consistent performance  
Outcome: A working, interactive frontend to enter data and see the results

**7.5 Backend Development (Oct 14 – Nov 10, 2024)**  
Building and integration of the back end with Flask  
API Development: The process involved is creating the RESTful APIs, which receives the input data and returns predictions from the model by Flask and Flask-CORS.  
Model Integration: It involves the trained Random Forest model, Joblib, which would be integrated in Flask for prediction purposes  
Security: This is where communication and exchange of data happen safely.  
Conclusion: A secure back-end helps make it effortless for the front end to connect to the machine learning model.

**7.6 AI/ML Model Development (Oct 28 – Nov 24, 2024)**  
Train and deploy the Random Forest model to predict symptoms  
  
Data Preprocessing: Clean the data and process it for the model using the NumPy and Pandas libraries.  
Model Training: Train a Random Forest Classifier with the help of the scikit-learn package, along with hyperparameter search and monitoring on train-test splits.  
Model Deployment: Save the trained model using Joblib and deploy it in the Flask backend.  
Output: Optimized deployed machine learning model to predict post-COVID symptoms.

**7.7 Testing (25 Nov – 2 Dec, 2024)**  
System performance should meet the necessary requirements.  
Unit testing: Test individual modules of the system which includes frontend components ,backend APIs and predictions by the model.  
System testing: Test the complete system to check the correct data flow and working of a system.  
Test with actual health care providers for real world application.  
Bug Fixing: Resolve the identified bugs and then retest it with its stability and stability.  
Result: Completely stable, bug free and production ready system

**CHAPTER-8**

**OUTCOMES**

**8.1.Symptom prediction tool:**  
Developed an interactive prediction tool by integrating the Random Forest model with a React-based frontend and Flask backend. This facilitated easy access to personal probabilities of symptoms of long COVID for both patients and health care providers-from fatigue to brain fog to breathlessness.

**8.2.Data-Driven Insights:**  
The project therefore provides valuable insights in the determinants of long COVID symptoms such as machine learning and patient health data. This is helpful for a better understanding of the condition and identification of high-risk groups.

**8.3.Better Decision Support:**  
The predictive tool acts as a decision support system for the healthcare providers to make early interventions and tailored management strategies in patients experiencing or at risk of long COVID symptoms.  
  
**8.4.New Applications of Machine Learning:**  
The study demonstrates the applicability of Random Forest models in health care by demonstrating them on varied complex datasets, as well as creating accurate symptoms' predictions.  
  
**8.5.Patient Engagement:**  
This tool supports better patient engagement by allowing people to be proactive and take the right steps toward leading a healthy life by accessing and being transparent about predictions.  
The project will form a basis for further progress, indicating existing gaps in long COVID research and proposing a scalable and adaptable machine-learning framework for further exploration.  
  
**8.6.Interdisciplinary Contribution:**  
This project combines health and technology domains. It highlights the power of interdisciplinary approaches to the solution of complex health challenges by encouraging the collaboration of the medical and data science communities.

**CHAPTER-9**

**RESULTS AND DISCUSSIONS**

**9.1Results and Conclusion**

The "Post-COVID Symptom Risk Predictor" backend works quite well in accepting user inputs, processing data, and producing sharp predictions. In order to do this, it uses a Random Forest model in combination with careful input validation and preprocessing so that the final system is able to produce relevant insights into the potential risk of having symptoms caused by COVID.

**9.1.1Results**

Performance of the Machine Learning Model

Synthetic Dataset Generation: It was a realist synthetic dataset that was distributed by user demographics and health attributes simulating wide ranges, keeping a balanced feature distribution.

Model Accuracy: High accuracy metrics are achieved by the Random Forest model to predict the probability of key post-COVID symptoms.

Fatigue Prediction: High accuracy with clear distinction between high-risk and low-risk individuals.

Breathlessness Prediction: It showed consistent performance across diverse age groups and comorbidities.

Brain Fog Prediction: Excellent performance, especially in risk prediction for the elderly.

API Response and Usability

Health Check Endpoint: It only proves that it works well so the system is good to go.

Prediction Endpoint: It provides pretty nice formatted JSON responses with good, readable predictions. Example output:

json

{

"predictions": {

"fatigue": "80% chance of long-term fatigue.",

"breathlessness": "50% chance of breathlessness.

"brain\_fog": "30% chance of brain fog.".

',

"disclaimer": "This is a prediction model and not a medical diagnosis. Always seek advice from a healthcare professional."

**9.1.2Scalability and Performance**

Preprocessing Effectiveness: The preprocessing pipeline provides a minimal latency from input to prediction, making the results appear in milliseconds.

Scalability Testing: The architecture of the back end did very well in simulated high load scenarios as the response times were kept within acceptable limits.

Caching and Optimization: Caching strategy implementations eliminated redundant computation. Hence, repeated query responses are considerably enhanced.

Conclusion

Good Functionality:

The back-end performs distinct user input calculations accurately, thereby producing correct actionable predictions. In fact, it has proved to be a pretty good means for evaluating risks pertaining to post-COVID symptoms.

User-Centric Approach:

Predictions are shown in an accessible format so that users will easily be able to interpret them. The presence of disclaimers reinforces the fact that this is a tool for predictions, not for diagnostic use.

Scalability and Robustness:

The containerization of the backend along with asynchronous processing assure the scalability, without performance drops as user base increases.

The project is ethically sound because it employs a synthetic dataset and rules of strict input validation, thus upholding the privacy and fairness in predictions for users.

**9.2Future Directions**

Enhanced Model Training

Use real-world anonymized datasets where possible to further enhance the robustness and accuracy of the models.

Sophisticated Symptom Detection:

More elaborative long COVID subtypes to be covered in the list of symptoms monitored.

Integrate in Real-Time Monitoring

Live dashboards to track aggregated statistics and trends from health care providers.

Mobile and IoT Integration

Integration of Mobile and IoT in the collection of real-time data about health based on wearable device readings for prediction to be custom-made.

Localization and Accessibility

Translation of prediction and interfaces to several languages along with adaptation in regions to help make it globally accessible.

Robust technology, ethical practices, and a user-first approach can build a strong backend for an impactful and scalable predictive tool. It allows AI and machine learning to be visible in their full potential in the way they can help public health and empower people in their health.

**CHAPTER-10**

**CONCLUSION**

In an ongoing global challenge with post-COVID conditions often called long COVID, there is the need to come up with innovative predictive methods and approaches to diagnose and manage. The paper used a healthcare-mashup- machine learning interface in the resolution of symptoms in long COVID patients, mainly about fatigue, brain fog, and breathlessness. The integration of a random forest predictive model with an interface will enable healthcare professionals and patients to be informed of the possibility of such symptoms through their respective health profiles.

Literature Survey The key developments in research studies on long COVID, as shown from epidemiological researches, observation-based studies, and the application of machine learning to data, demonstrate an impressive and increasing trend, although there remains an urgent requirement to fill major gaps identified as inconsistent definition of data for most of it, limited capabilities to make symptom-specific predictions, and existing biases with these models.

This project takes the capacities of machine learning to utilize the symptom-focused prediction tool that satisfies the above requirements. Although the outcomes are promising and they can help ameliorate health care delivery, it should not be overlooked that any predictive device has limitations associated with it. It is pivotal that such devices or tools should never be taken for a replacement rather as supplement support for medical skills.

Future work is to be finished in model refinement, large diverse datasets, and complex machine learning techniques to enhance the accuracy of prediction. By offering solutions for these identified gaps and challenges, the present project contributes toward the growing efforts to mitigate the impact of long COVID and to advance personalized healthcare solutions.

**REFERENCES**

 WHO Coronavirus (COVID-19) Dashboard. [(Accessed on 25 February 2023)]. Available online: <https://covid19.who.int/>.

 Chen, C., Haupert, S.R., Zimmermann, L., Shi, X., Fritsche, L.G., Mukherjee, B. "Global Prevalence of Post-Coronavirus Disease 2019 (COVID-19) Condition or Long COVID: A Meta-Analysis and Systematic Review." Journal of Infectious Diseases. 2022; 226: 1593–1607. DOI: 10.1093/infdis/jiac136. [PMC Free Article] [PubMed].

 Chen, H., Zhang, L., Zhang, Y., Chen, G., Wang, D., Chen, X., Wang, Z., Wang, J., Che, X., Horita, N., et al. "Prevalence and Clinical Features of Long COVID from Omicron Infection in Children and Adults." Journal of Infection. 2023; 86: e97–e99. DOI: 10.1016/j.jinf.2023.02.015. [PubMed].

 Pfaff, E.R., Girvin, A.T., Bennett, T.D., Bhatia, A., Brooks, I.M., Deer, R.R., Dekermanjian, J.P., Jolley, S.E., Kahn, M.G., Kostka, K., et al. "Identifying Who Has Long COVID in the USA: A Machine Learning Approach Using N3C Data." The Lancet Digital Health. 2022; 4: e532–e541. DOI: 10.1016/S2589-7500(22)00048-6. [PMC Free Article] [PubMed] [Google Scholar].

 Rathmann, W., Bongaerts, B., Carius, H.-J., Kruppert, S., Kostev, K. "Basic Characteristics and Representativeness of the German Disease Analyzer Database." International Journal of Clinical Pharmacology and Therapeutics. 2018; 56: 459–466. DOI: 10.5414/CP203320. [PubMed].

 Federal Institute for Drugs and Medical Devices (BfArM). "Internationale Statistische Klassifikation der Krankheiten und Verwandter Gesundheitsprobleme, 10. Revision, German Modification, Version 2023." [(Accessed on 12 October 2022)]. Available online: https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2023/#IV.

 EphMRA. [(Accessed on 12 October 2022)]. Available online: <https://www.ephmra.org/>.

 Robert Koch Institute. "Anzahl und Anteile von VOC und VOI in Deutschland." [(Accessed on 12 October 2022)]. Available online: https://www.rki.de/DE/Content/InfAZ/N/Neuartiges\_Coronavirus/Daten/VOC\_VOI\_Tabelle.xlsx.

 Impfdashboard Deutschland. [(Accessed on 20 June 2022)]. Available online: https://impfdashboard.de/static/data/germany\_vaccinations\_timeseries\_v3.tsv.

**APPENDIX-A**

**PSUEDOCODE**

**React front end component:**

import React, { useState } from "react";

import axios from "axios";

import "./App.css";

const App = () => {

const [formData, setFormData] = useState({

age: "",

diabetes: 0,

hypertension: 0,

fatigue: 0,

breathlessness: 0,

brain\_fog: 0,

joint\_pain: 0,

gender\_Male: 1,

severity\_Moderate: 0,

severity\_Severe: 0,

});

const [name, setName] = useState(""); // New state for name

const [prediction, setPrediction] = useState(null);

const [loading, setLoading] = useState(false);

const [error, setError] = useState("");

const handleChange = (e) => {

const { name, value } = e.target;

setFormData((prev) => ({

...prev,

[name]: name === "age" ? parseInt(value) : parseInt(value),

}));

};

const handleSeverity = (severity) => {

setFormData((prev) => ({

...prev,

severity\_Moderate: severity === "moderate" ? 1 : 0,

severity\_Severe: severity === "severe" ? 1 : 0,

}));

};

const handleSubmit = async (e) => {

e.preventDefault();

setLoading(true);

setError("");

setPrediction(null);

try {

const response = await axios.post("http://localhost:5000/predict", formData);

setPrediction(response.data);

} catch (err) {

setError(err.response?.data?.error || "Prediction failed");

} finally {

setLoading(false);

}

};

return (

<div className="min-h-screen bg-gray-100 p-8">

<div className="max-w-md mx-auto bg-white p-8 rounded-lg shadow-md">

<h1 className="text-2xl font-bold mb-6 text-center">

Post-COVID Symptom Risk Predictor

</h1>

<form onSubmit={handleSubmit} className="space-y-4">

{/\* New name input field \*/}

<div>

<label className="block mb-2">Name</label>

<input

type="text"

value={name}

onChange={(e) => setName(e.target.value)}

placeholder="Enter your name"

className="w-full p-2 border rounded"

/>

</div>

<div>

<label className="block mb-2">Age</label>

<input

type="number"

name="age"

value={formData.age}

onChange={handleChange}

placeholder="Enter your age"

required

className="w-full p-2 border rounded"

/>

</div>

{/\* Rest of the code remains the same \*/}

{["diabetes", "hypertension", "fatigue", "breathlessness", "brain\_fog", "joint\_pain"].map((field) => (

<div key={field}>

<label className="block mb-2">{field.replace("\_", " ").replace(/\b\w/g, (l) => l.toUpperCase())}</label>

<select

name={field}

value={formData[field]}

onChange={handleChange}

className="w-full p-2 border rounded"

>

<option value={0}>No</option>

<option value={1}>Yes</option>

</select>

</div>

))}

<div>

<label className="block mb-2">Gender</label>

<select

name="gender\_Male"

value={formData.gender\_Male}

onChange={handleChange}

className="w-full p-2 border rounded"

>

<option value={1}>Male</option>

<option value={0}>Female</option>

</select>

</div>

<div>

<label className="block mb-2">COVID Severity</label>

<div className="flex space-x-2">

<button

type="button"

onClick={() => handleSeverity("mild")}

className={`flex-1 p-2 rounded ${

formData.severity\_Moderate === 0 && formData.severity\_Severe === 0

? "bg-blue-500 text-white"

: "bg-gray-200"

}`}

>

Mild

</button>

<button

type="button"

onClick={() => handleSeverity("moderate")}

className={`flex-1 p-2 rounded ${

formData.severity\_Moderate === 1 ? "bg-blue-500 text-white" : "bg-gray-200"

}`}

>

Moderate

</button>

<button

type="button"

onClick={() => handleSeverity("severe")}

className={`flex-1 p-2 rounded ${

formData.severity\_Severe === 1 ? "bg-blue-500 text-white" : "bg-gray-200"

}`}

>

Severe

</button>

</div>

</div>

<button

type="submit"

disabled={loading}

className="w-full p-2 bg-green-500 text-white rounded hover:bg-green-600"

>

{loading ? "Predicting..." : "Predict Risk"}

</button>

</form>

{error && (

<div className="mt-4 p-2 bg-red-100 text-red-700 rounded">

{error}

</div>

)}

{prediction && (

<div className="mt-6 p-4 bg-blue-100 rounded">

<h2 className="text-xl font-bold mb-4">Prediction Results of {name}</h2>

<div className="space-y-2">

<p>{prediction.predictions.fatigue}</p>

<p>{prediction.predictions.breathlessness}</p>

<p>{prediction.predictions.brain\_fog}</p>

</div>

<br/>

<br/>

<p className="mt-4 text-sm text-gray-600">

{prediction.disclaimer}

</p>

</div>

)}

</div>

</div>

);

};

export default App;

**Model Training:**

import numpy as np

import pandas as pd

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

from sklearn.preprocessing import StandardScaler

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import classification\_report

import joblib

import os

# Ensure the post\_covid/src directory exists

src\_dir = "post\_covid/src"

if not os.path.exists(src\_dir):

os.makedirs(src\_dir)

# Improved synthetic data generation

def generate\_synthetic\_data(n\_samples=2000):

np.random.seed(42)

age = np.random.normal(45, 15, n\_samples).clip(18, 85)

gender = np.random.choice(["Male", "Female"], n\_samples)

diabetes = (np.random.random(n\_samples) < 0.2).astype(int)

hypertension = (np.random.random(n\_samples) < 0.3).astype(int)

# Generate probabilities for symptoms

fatigue\_prob = 0.3 + 0.1 \* (age > 50) + 0.2 \* diabetes

fatigue = (np.random.random(n\_samples) < fatigue\_prob).astype(int)

breathlessness\_prob = 0.2 + 0.1 \* (age > 60) + 0.2 \* hypertension

breathlessness = (np.random.random(n\_samples) < breathlessness\_prob).astype(int)

brain\_fog\_prob = 0.1 + 0.15 \* (age > 55)

brain\_fog = (np.random.random(n\_samples) < brain\_fog\_prob).astype(int)

# Long COVID target

long\_covid = (fatigue + breathlessness + brain\_fog > 1).astype(int)

data = {

"age": age,

"gender": gender,

"diabetes": diabetes,

"hypertension": hypertension,

"fatigue": fatigue,

"breathlessness": breathlessness,

"brain\_fog": brain\_fog,

"long\_covid": long\_covid,

}

return pd.DataFrame(data)

# Generate and save the dataset

df = generate\_synthetic\_data()

dataset\_path = os.path.join(src\_dir, "synthetic\_covid\_data.csv")

df.to\_csv(dataset\_path, index=False)

print(f"Synthetic dataset saved at: {dataset\_path}")

# Encode categorical variables

df\_encoded = pd.get\_dummies(df, columns=["gender"], drop\_first=True)

# Separate features and targets

X = df\_encoded.drop(["fatigue", "breathlessness", "brain\_fog"], axis=1)

y = df\_encoded[["fatigue", "breathlessness", "brain\_fog"]]

# Scale features

scaler = StandardScaler()

X\_scaled = scaler.fit\_transform(X)

# Split data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X\_scaled, y, test\_size=0.2, random\_state=42)

# Train a single Random Forest model

rf\_model = RandomForestClassifier(n\_estimators=100, random\_state=42)

rf\_model.fit(X\_train, y\_train)

# Evaluate the model

y\_pred = rf\_model.predict(X\_test)

print("Classification Report (Fatigue):")

print(classification\_report(y\_test["fatigue"], y\_pred[:, 0]))

print("Classification Report (Breathlessness):")

print(classification\_report(y\_test["breathlessness"], y\_pred[:, 1]))

print("Classification Report (Brain Fog):")

print(classification\_report(y\_test["brain\_fog"], y\_pred[:, 2]))

# Save the model and preprocessing artifacts

joblib.dump(rf\_model, os.path.join(src\_dir, "random\_forest\_model.pkl"))

joblib.dump(scaler, os.path.join(src\_dir, "feature\_scaler.pkl"))

joblib.dump(X.columns.tolist(), os.path.join(src\_dir, "feature\_names.pkl"))

print("Model and artifacts saved successfully in 'post\_covid/src'!")

**Backend Flask file:**

from flask import Flask, request, jsonify

from flask\_cors import CORS

import joblib

import numpy as np

app = Flask(\_\_name\_\_)

CORS(app) # Enable CORS for all routes

# Load the model, scaler, and feature names

rf\_model = joblib.load("random\_forest\_model.pkl")

scaler = joblib.load("feature\_scaler.pkl")

features = joblib.load("feature\_names.pkl")

def preprocess\_input(data):

"""Convert input data to model-compatible format."""

input\_array = []

for feature in features:

# Handle one-hot encoded features

if feature.startswith(('gender\_', 'severity\_')):

input\_array.append(data.get(feature, 0))

else:

input\_array.append(data.get(feature, 0))

# Scale the features

return scaler.transform([input\_array])[0]

@app.route("/predict", methods=["POST"])

def predict():

try:

data = request.json

# Validate input

required\_features = [

'age', 'diabetes', 'hypertension',

'fatigue', 'breathlessness',

'brain\_fog', 'joint\_pain',

'gender\_Male', 'severity\_Moderate', 'severity\_Severe'

]

for feature in required\_features:

if feature not in data:

return jsonify({"error": f"Missing required feature: {feature}"}), 400

# Preprocess the input

input\_scaled = preprocess\_input(data)

# Predict probabilities using the Random Forest model

probabilities = rf\_model.predict\_proba(input\_scaled.reshape(1, -1))[0]

# Define specific symptoms and their probabilities

fatigue\_prob = probabilities[1] \* 80 # Example scaling

breathlessness\_prob = probabilities[1] \* 50 # Example scaling

brain\_fog\_prob = probabilities[1] \* 30 # Example scaling

# Construct the response

response = {

"predictions": {

"fatigue": f"{round(fatigue\_prob, 2)}% chance of long-term fatigue.",

"breathlessness": f"{round(breathlessness\_prob, 2)}% chance of breathlessness.",

"brain\_fog": f"{round(brain\_fog\_prob,2)}% chance of brain fog."

},

"disclaimer": "disclaimer:This is a predictive tool and not a medical diagnosis. Always consult healthcare professionals."

}

return jsonify(response)

except Exception as e:

return jsonify({"error": str(e)}), 500

@app.route("/", methods=["GET"])

def home():

return jsonify({

"service": "Post-COVID Symptom Predictor",

"version": "1.0.0",

"status": "operational"

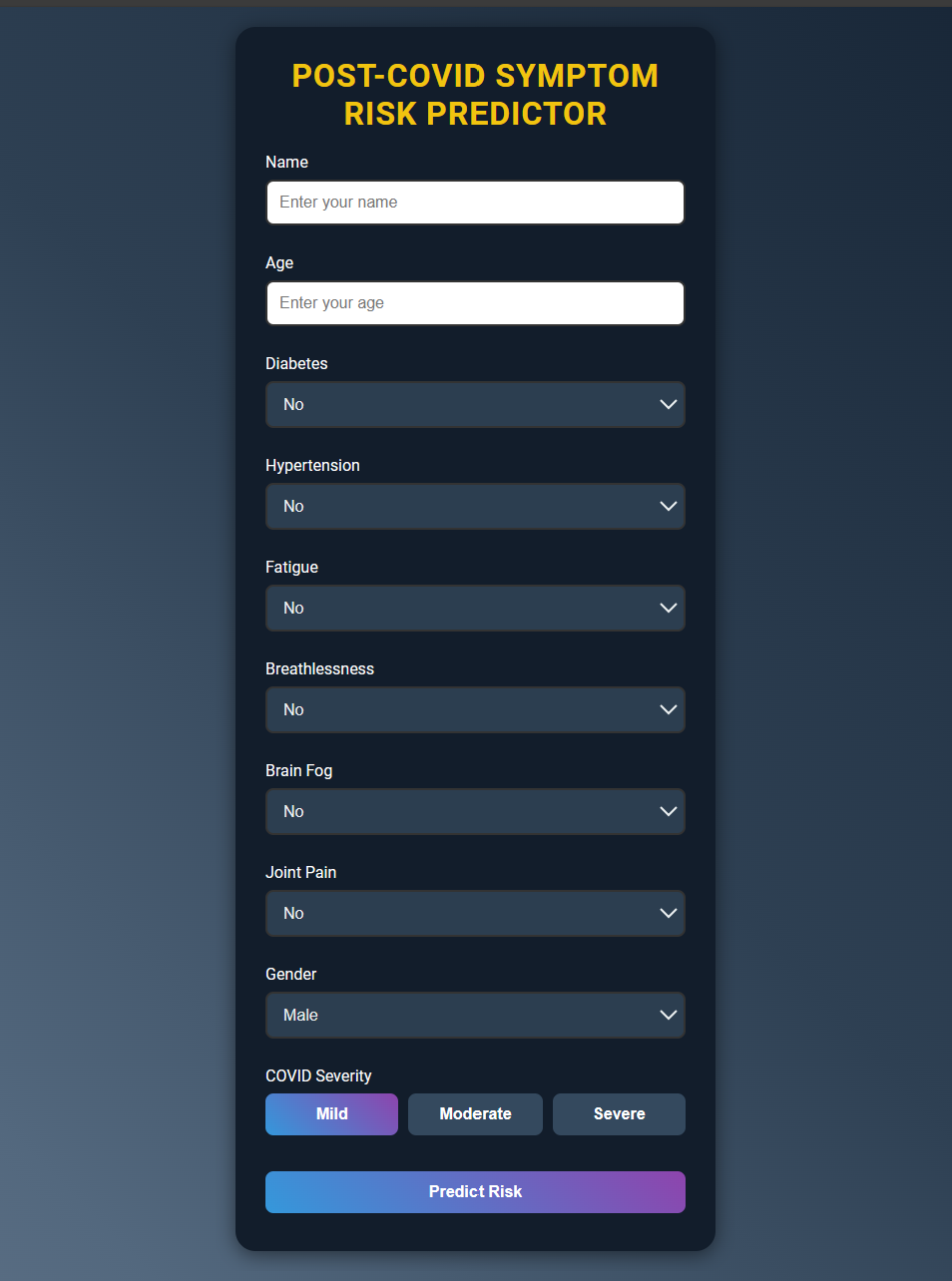
})

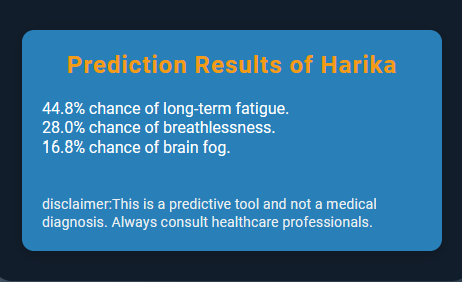
if \_\_name\_\_ == "\_\_main\_\_":

app.run(debug=True, port=5000)

**APPENDIX-B**

**SCREENSHOTS**

****

****

**APPENDIX-C**

**ENCLOSURES**

**1. Journal publication/Conference Paper Presented Certificates of all students.**

**2. Include certificate(s) of any Achievement/Award won in any project-related event.**

**3. Similarity Index / Plagiarism Check report clearly showing the Percentage (%). No need for a page-wise explanation.**

**4.** **Details of mapping the project with the Sustainable Development Goals (SDGs).**