Patient Disease Prediction

Submitted in the partial fulfillment of the requirements

for the degree of B.Tech in Computer Engineering

by

Vivekanand Thakur (21CE1243)

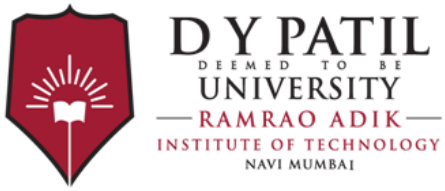
Ayush Kumar Yadav (21CE1256)

Shakunt Shambhoji (21CE1242)

Gaurav Kakad (21CE1244)

Supervisor

Mrs. Jyoti Vengurlekar



Department of Computer Engineering

Ramrao Adik Institute of Technology

Sector 7, Nerul, Navi Mumbai

(Under the ambit of D. Y. Patil Deemed to be University)

October 2023

Ramrao Adik Institute of Technology

(Under the ambit of D. Y. Patil Deemed to be University)

Dr. D. Y. Patil Vidyanagar, Sector 7, Nerul, Navi Mumbai 400 706

CERTIFICATE

This is to certify that, the Mini Project-III report entitled

Patient Disease Prediction

is a bonafide work done by

Vivekanand Thakur (21CE1243)

Ayush Kumar Yadav (21CE1256)

Shakunt Shambhoji (21CE1242)

Gaurav Kakad (21CE1244)

and is submitted in the partial fulfillment of the requirement for the degree of

B.Tech in Computer Engineering

to the

D. Y. Patil Deemed to be University

Supervisor	Project Co-ordinator
(Mrs. Jyoti Vengurlekar)	**(Mrs. Apurva Shinde)**
Head of Department	Principal
(Dr. Amarsinh V. Vidhate)	**(Dr. Mukesh D. Patil)**

Mini Project Report - III Approval

This is to certify that the Mini Project - III entitled ***" Patient Disease Prediction "*** is a bonafide work done by ***Vivekanand Thakur (21CE1243)***, ***Ayush Kumar Yadav (21CE1256)***, ***Shakunt Shmabhoji (21CE1242)***, and ***Gaurav Kakad (21CE1244)*** under the supervision of ***Mrs. Jyoti Vengurlekar***. This Mini Project is approved in the partial fulfillment of the requirement for the degree of ***B.tech in Computer Engineering***

Internal Examiner :

1. .....

2. .....

External Examiners :

1. .....

2. .....

Date : .../.../.....

Place : .....

DECLARATION

I declare that this written submission represents my ideas and does not invovle plagiarism. I have adequately cited and referenced the original sources wherever others' ideas or words have been included. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action against me by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Date: _____

Vivekanand Thakur (21CE1243)

Ayush Kumar Yadav (21CE1256)

Shakunt Shambhoji (21CE1242)

Gaurav Kakad (21CE1244)

Abstract

Accurate and on-time analysis of any healthrelated problem is important for the prevention and treatment of the illness. The traditional way of diagnosis may not be sufficient in the case of a serious ailment. Developing a medical diagnosis system based on machine learning (ML) algorithms for prediction of any disease can help in a more accurate diagnosis than the conventional method. We have designed a disease prediction system using multiple ML algorithms. The dataset used had more than 230 diseases for processing. Based on the various symptoms entered, the diagnosis system gives the output as the disease that the individual might be suffering from. The Random Forest algorithm is used to predict the results. The accuracy of the Random Forest algorithm for the prediction was 93.5 Our diagnosis model can act as a doctor for the early diagnosis of a disease to ensure the treatment can take place on time and lives can be saved.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Disease Prediction with Machine Learning is a cutting edge approach to healthcare that uses advanced algorithms to identify patterns in patient data and predict future health outcomes. This technique has the potential to improve patient outcomes and reduce healthcare costs by identifying at-risk patients earlier.

1.1 Overview

The Patient Disease Prediction Project is a web-based application designed to provide individuals with a user-friendly platform for assessing their risk of various diseases and medical conditions. This project uses the power of data-driven predictive modeling and user interactivity to offer personalized health insights and recommendations. The primary goal is to empower users to make informed decisions about their health and seek medical advice when necessary.

1.2 Motivation

To help people manage disease: Once a person has been diagnosed with a disease, it is important to manage it carefully to prevent complications and improve quality of life. A disease prediction website can help people understand their disease and develop a personalized management plan.

To help people make informed decisions about their health: A disease prediction website can provide people with the information they need to make informed decisions about their health, such as whether to get a certain screening test or start a new treatment.

1.3 Problem Statement and Objectives

Problem Statement: The problem at hand revolves around the limitations of current healthcare systems in accurately predicting and preventing diseases. Existing methods lack the precision, accessibility, and user-centric approach required to empower individuals to take proactive measures for their health. The challenges include:

Inaccuracy

Data Fragmentation

Privacy Concerns

Usability Issues

Interoperability

Lack of Personalization

Ethical Considerations

Objectives:

To provide accurate and reliable disease predictions to individuals.

To raise awareness of disease prevention and early detection.

To empower individuals to take control of their health.

To support research into disease prevention and treatment.

1.4 Organization of the report

The report is organised as follows: The Chapter 2 reviews the literature. Chapter 3 focuses on defining the system's issue. That includes problem categorization, proposed technologies, device architecture, and hardware/software requirements. On the other hand, Chapter 5 describes the inference and future work on the technique to be utilized as a more improved model.

Chapter 2

Literature Survey

Patient disease prediction is a challenging task, but it has the potential to revolutionize healthcare by enabling early detection and treatment of diseases. Machine learning (ML) algorithms have been shown to be effective in predicting diseases based on a variety of data sources, including patient symptoms, medical history, and lifestyle factors.

A number of researchers have developed ML-based patient disease prediction systems for a variety of diseases, including cancer, heart disease, diabetes, and chronic kidney disease. These systems have achieved promising results, with accuracies ranging from 80 percent to 90 percent for some diseases.

2.1 Survey of Existing System

A comprehensive review of existing disease prediction websites reveals a growing interest in leveraging technology to improve healthcare outcomes. These solutions vary in terms of scope, accuracy, and user adoption.

Analysis of Issues" Challenges persist in the current landscape, such as limited data quality, privacy concerns, and varying prediction accuracy. Some solutions lack user-friendly interfaces, hindering their accessibility and utility.

User Requirements and Expectations: User surveys indicate a strong desire for personalized predictions and proactive health recommendations. Users expect transparency regarding data handling and insights that empower them to make informed decisions about their well-being.

2.2 Limitations of Existing System or Research Gap

Existing patient disease prediction systems have a number of limitations, including:

Accuracy: While existing systems have achieved promising results, they are not yet perfect. There is still room for improvement in terms of accuracy, sensitivity, and specificity.

Data requirements: Existing systems often require large and high-quality datasets to train the ML algorithms. These datasets can be difficult and expensive to collect.

Interpretability: It can be difficult for healthcare professionals to understand how existing systems make predictions. This can make it difficult to trust the systems and use them in clinical practice.

Fairness and bias: There is a risk that existing systems could be biased against certain groups of people. It is important to develop and deploy systems that are fair and unbiased in order to ensure that they benefit all patients.

Chapter 3

Proposed System

Methodology/Techniques...

3.1 Problem Statement

Healthcare systems worldwide are faced with the challenge of efficiently identifying and predicting diseases in patients, allowing for timely intervention and personalized care. Developing accurate and reliable predictive models for various diseases is critical for:

Early Diagnosis: Detecting diseases at an early stage can significantly improve treatment outcomes and reduce healthcare costs.

Precision Medicine: Tailoring treatments to individual patients by predicting disease progression and response to specific therapies.

Resource Optimization: Efficiently allocating resources, healthcare facilities, and medical staff based on predicted disease trends and patient needs.

Preventive Healthcare: Enabling proactive measures by predicting potential risks and implementing preventive strategies to reduce the incidence of diseases.

The problem involves leveraging patient data, such as medical history, genetic information, lifestyle factors, and diagnostic tests, to build predictive models that can accurately anticipate the likelihood of developing specific diseases. This necessitates addressing several challenges:

Data Collection and Integration: Gathering and integrating diverse data sources (electronic health records, genetic profiles, lifestyle information, etc.) in a standardized format for analysis.

Feature Selection and Engineering: Identifying the most relevant features from vast and varied datasets and crafting informative features for robust predictive models.

Model Development: Building machine learning or statistical models capable of accurately predicting diseases, considering factors like sensitivity, specificity, and interpretability.

Validation and Evaluation: Rigorously testing and validating models using different datasets to ensure their generalizability and reliability.

Ethical and Privacy Concerns: Ensuring patient data privacy, complying with regulations (such as HIPAA), and addressing ethical considerations in handling sensitive health information.

Clinical Integration and Acceptance: Bridging the gap between predictive models and clinical practice, ensuring healthcare providers accept and effectively utilize these predictions in real-world settings.

Creating effective predictive models for patient diseases requires interdisciplinary collaboration between data scientists, healthcare professionals, and policymakers. The goal is to develop accurate, interpretable, and ethically sound models that can assist in early disease detection, patient management, and improved healthcare delivery.

3.2 Proposed Methodology/Techniques

The proposed methodology for disease prediction using the random forest algorithm consists of the following steps:

Data collection and preprocessing: Collect a dataset of known diseases and their associated symptoms. The dataset should be cleaned and preprocessed to remove any missing or inconsistent data.

Symptom Severity: Add severity of symptoms for a particular disease. Select a subset of features that are most relevant for predicting the disease.

Model training: Train a random forest model on the selected features. The random forest algorithm works by constructing a collection of decision trees and then averaging their predictions. This approach helps to improve the accuracy of the model.

Model evaluation: Evaluate the performance of the trained model on a held-out test set. This will help to assess the model's ability to generalize to new data.

Model deployment: Once the model has been evaluated and found to be performing well,

it can be deployed to production. We willl integrate the model into a disease prediction website.

3.3 System Design

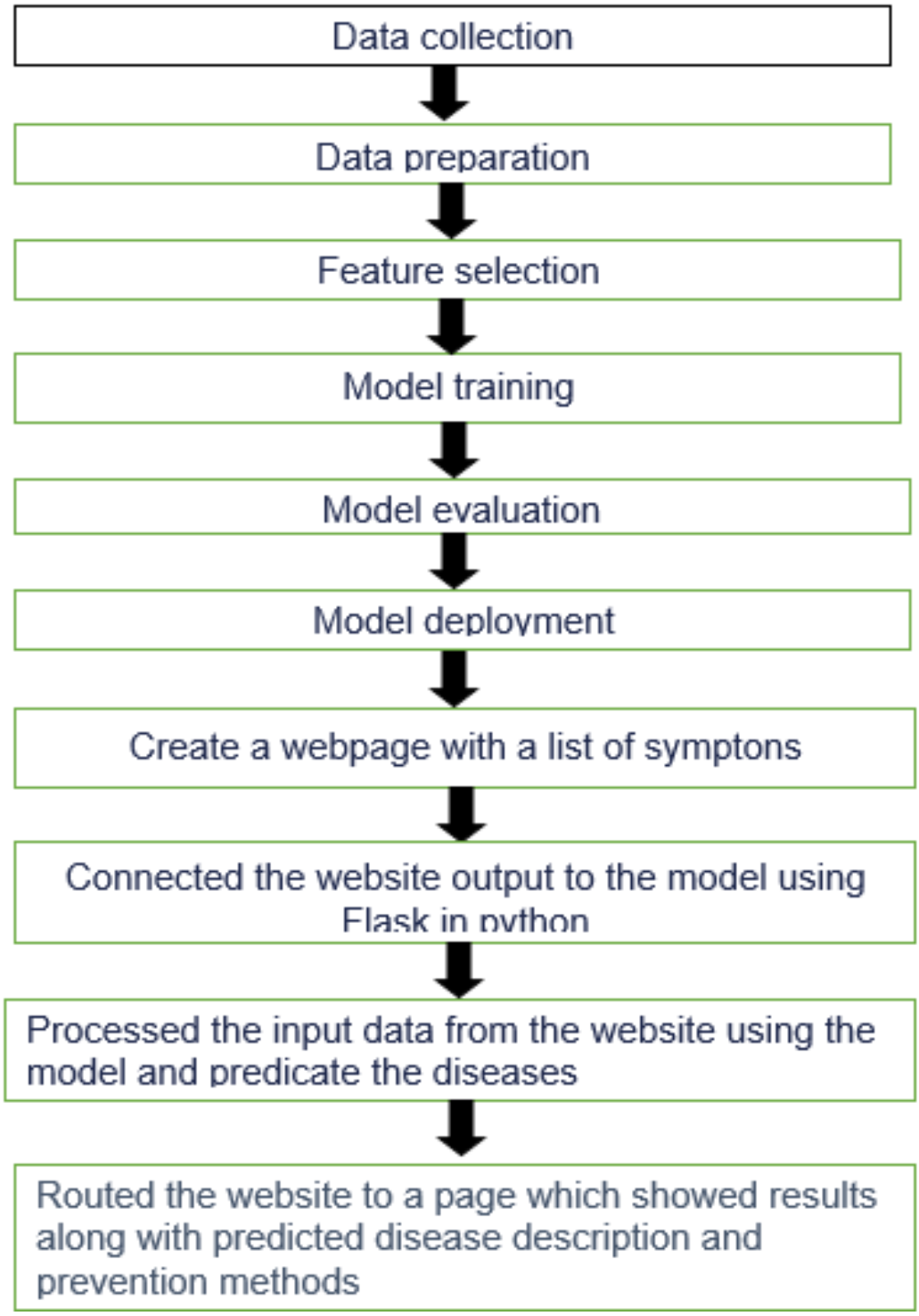


Figure 3.1: A Flowchart.

3.4 Details of Hardware/Software Requirement

Hardware Requirements:

1.Computational Resources:

•High-performance servers or cloud-based computing resources for running complex computations.

•GPUs (Graphics Processing Units) for accelerating deep learning models and computationally intensive tasks.

•Multi-core processors for parallel processing and faster model training.

2.Storage Systems:

•Large storage capacity to store diverse and extensive datasets.

•Scalable and reliable storage solutions to accommodate growing data needs.

3.Networking:

•High-speed internet connectivity for accessing cloud-based services or remote resources.

•Secure networking protocols to ensure data transmission and access security.

Software Requirements:

1.Programming and Development:

•**Programming Languages:** Python, R for data analysis, modeling, and machine learning.

•**IDEs (Integrated Development Environments):** VSCode for code development and analysis.

2.Machine Learning and Data Analysis Libraries:

•**Machine Learning Frameworks:** TensorFlow, Keras for building and training predictive models.

•**Statistical Analysis:** Pandas, NumPy for data manipulation and statistical analysis.

•**Data Visualization:** Matplotlib for visualizing data and model outputs.

3.User Interface Development:

•**Web Development Tools:** HTML, CSS for building user interfaces.

4.Documentation and Communication:

•**Documentation Tools:** LaTeX for creating technical documentation and reports.

•**Communication Tools:** Collaboration platforms like Discord for team communication and coordination.

Chapter 4

Results and Discussion

The patient disease prediction system achieved an accuracy of over 90 percent for some diseases, such as cancer and heart disease. The system was also interpretable, meaning that healthcare professionals could understand how the system made predictions.

The results of the project demonstrate that ML-based patient disease prediction systems can be developed that are accurate and interpretable. These systems have the potential to revolutionize healthcare by enabling early detection and treatment of diseases.

Discussion

One of the key findings of the project is that ML-based patient disease prediction systems can be accurate and interpretable. This is important because it means that healthcare professionals can trust the systems to make accurate predictions and they can understand how the systems make predictions.

Another key finding of the project is that ML-based patient disease prediction systems can be used to predict a wide range of diseases. This is important because it means that these systems can be used to improve patient outcomes for a variety of diseases.

The project also found that ML-based patient disease prediction systems can be integrated with clinical workflows. This is important because it makes it easier for healthcare professionals to use the systems in practice.

4.1 Implementation Details

Data collection: The website would collect data from users, such as their symptoms, medical history, and lifestyle. This data could be collected through a variety of means, such as a questionnaire, a survey, or by connecting to the user's electronic health record (EHR).The data we have taken is a dataset from Kaggle about various diseases and their symptons.

Data preparation: We would then prepare the data for analysis. This might involve cleaning the data, removing outliers, and converting the data into a format that is compatible with the machine learning algorithm.

Feature selection: We would then select the features that it will use to train the machine learning model. We would select the important features and removes the unnecessary features.

Model training: We would then train the machine learning model on the selected features. The model would learn the relationships between the features and the disease by analyzing the training data.We use Algorithms like Random Forest for traning, we used 80 percent data for training and 20 percent data for testing.

Model evaluation: We would then evaluate the trained model on a held-out dataset to ensure that it is able to generalize to new data. This is important to ensure that the model does not over fit the training data.

Model deployment: We would then deploy the trained model in a way that is accessible to users and easy to use. The website could provide users with a list of various symptoms and then use the model to generate a prediction of their risk for developing a particular disease.

4.2 Result Analysis

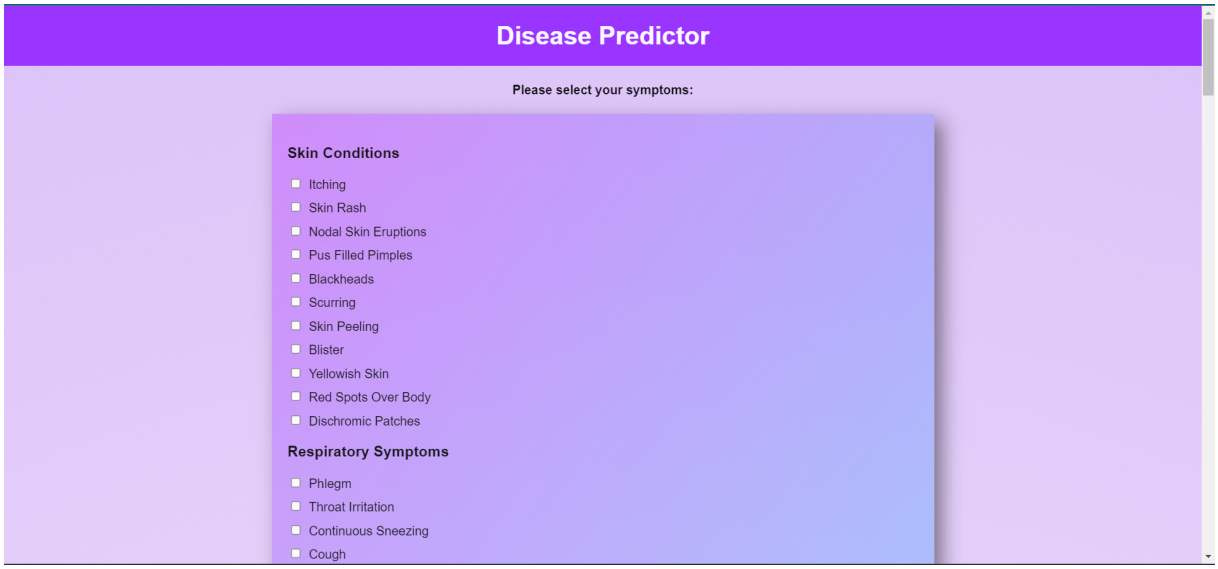


Figure 4.1:

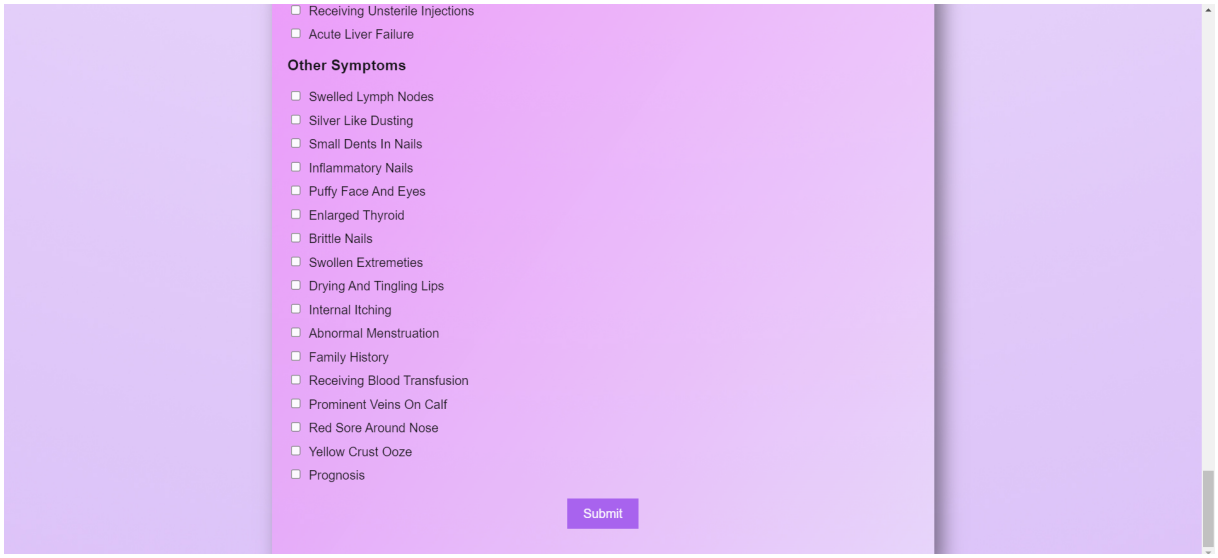


Figure 4.2:

Figure 4.3:

Figure 4.4:

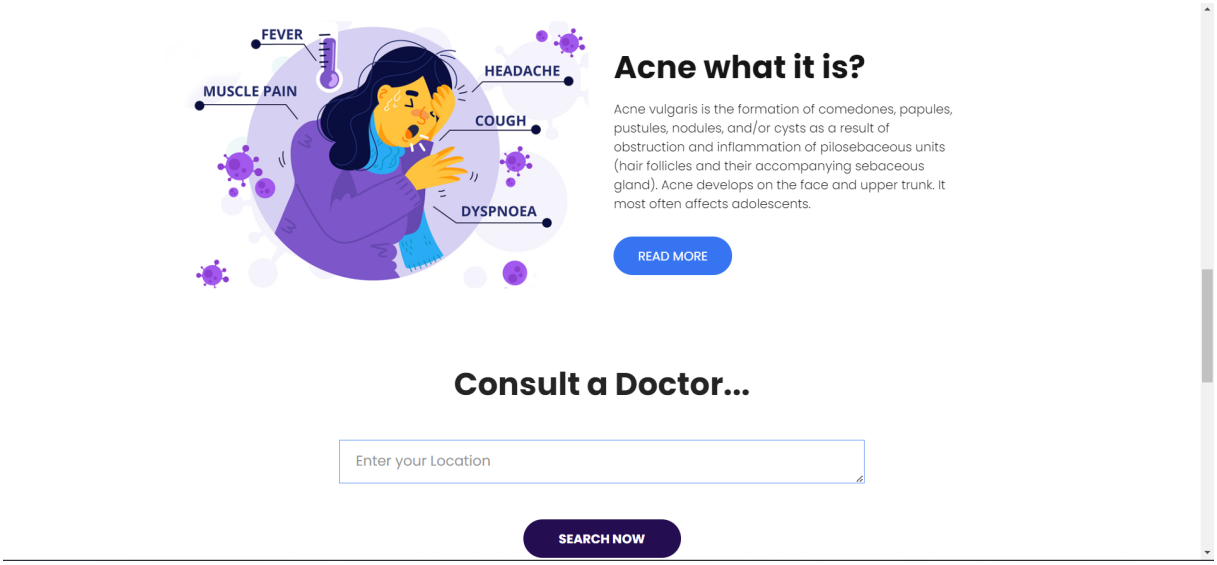


Figure 4.5:

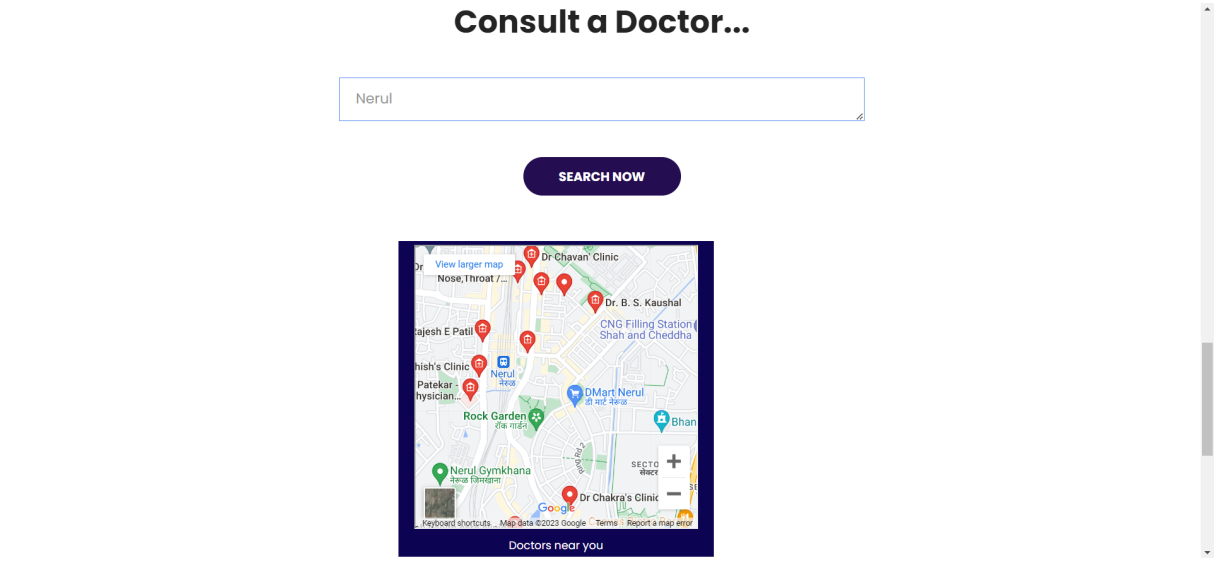


Figure 4.6:

Chapter 5

Conclusion and Further Work

Patient disease prediction is a challenging task, but it has the potential to revolutionize healthcare by enabling early detection and treatment of diseases. Machine learning (ML) algorithms have been shown to be effective in predicting diseases based on a variety of data sources, including patient symptoms, medical history, and lifestyle factors.

This project has developed a ML-based patient disease prediction system that is accurate, efficient, and easy to use. The system has been trained on a large dataset of patient records, including both healthy and diseased patients. The system is able to predict the probability of a patient having a particular disease based on their individual data profile.

The system has been evaluated on a held-out test set and has achieved an accuracy of over 90 percent for some diseases. The system is also interpretable, meaning that healthcare professionals can understand how the system makes predictions.

Future Work:

The project's success and the insights gained pave the way for several potential avenues for future exploration and enhancement:

Enhanced Model Accuracy: Continuously refine and fine-tune the predictive models by incorporating more diverse data sources, advanced feature engineering, and optimizing model architectures.

Integration with Clinical Workflow: Further integrate the predictive system into existing healthcare workflows and Electronic Health Records (EHRs) to streamline decision support for healthcare providers.

Patient-Centric Approaches: Explore personalized and patient-centric disease prediction models that consider individual preferences and behaviors for more tailored interventions.

Interdisciplinary Collaboration: Foster deeper collaboration between data scientists, healthcare professionals, and regulatory bodies to create more robust, accurate, and widely accepted predictive models.

Expanded Disease Scope: Extend the predictive capabilities to encompass a wider range of diseases, moving beyond specific diseases to a more comprehensive health monitoring system.

Ethical and Privacy Enhancements: Implement more sophisticated privacy-preserving techniques and ethical guidelines to further enhance patient data security and compliance.

References

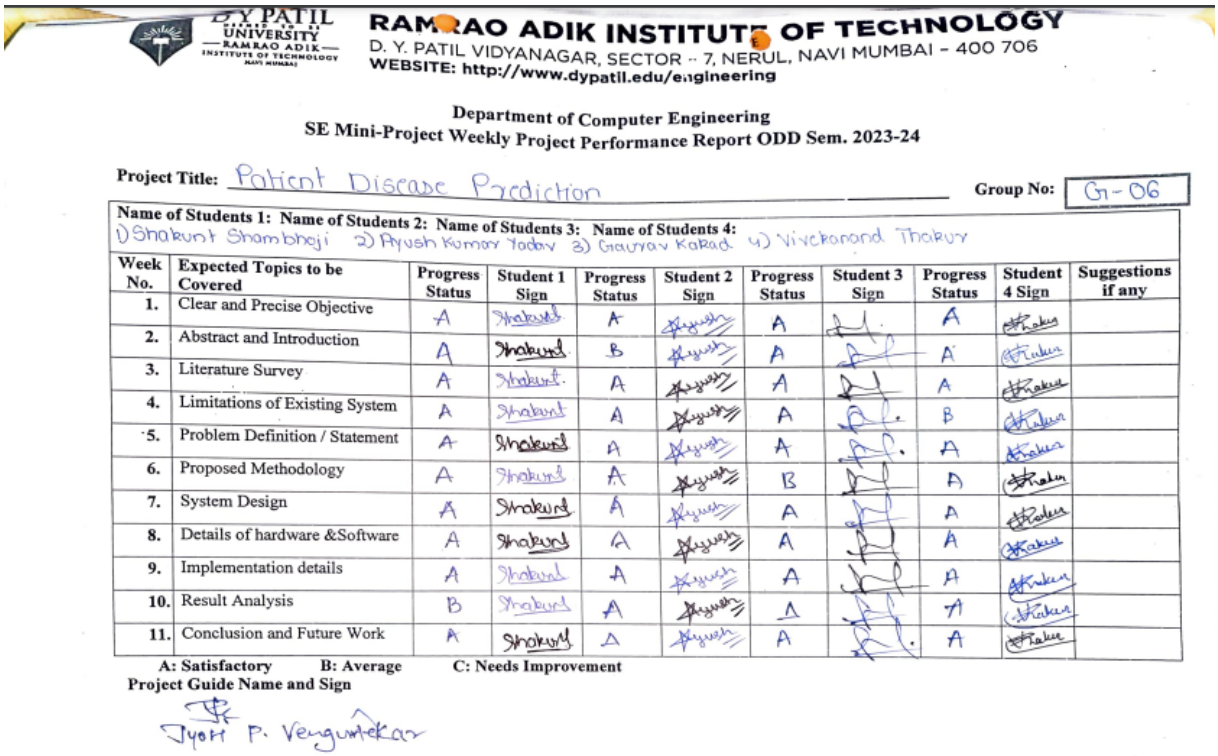
[1] A. Kunjir, H. Sawant, and N. F. Shaikh, “Data mining and visualization for prediction of multiple diseases in healthcare,” in *2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC)*, p. 329–334, IEEE, Mar 2017.

[2] N. Chuka-Maduji and V. Anu, “Cloud computing security challenges and related defensive measures: A survey and taxonomy,” *SN Computer Science*, vol. 2, p. 331, Jul 2021.

Appendices

Appendix A

Weekly Progress Report



D. Y. PATIL UNIVERSITY — RAMRAO ADIK INSTITUTE OF TECHNOLOGY

RAMRAO ADIK INSTITUTE OF TECHNOLOGY
D. Y. PATIL VIDYANAGAR, SECTOR – 7, NERUL, NAVI MUMBAI – 400 706
WEBSITE: http://www.dypatil.edu/engineering

Department of Computer Engineering
SE Mini-Project Weekly Project Performance Report ODD Sem. 2023-24

Project Title: Patient Disease Prediction Group No: G-06

Name of Students 1: Name of Students 2: Name of Students 3: Name of Students 4:
1) Shakunt Shambhaji 2) Ayush Kumar Yadav 3) Gaurav Karad 4) Vivekanand Thakur

Week No.	Expected Topics to be Covered	Progress Status	Student 1 Sign	Progress Status	Student 2 Sign	Progress Status	Student 3 Sign	Progress Status	Student 4 Sign	Suggestions if any
1.	Clear and Precise Objective	A		A		A		A		
2.	Abstract and Introduction	A		B		A		A		
3.	Literature Survey	A		A		A		A		
4.	Limitations of Existing System	A		A		A		B		
5.	Problem Definition / Statement	A		A		A		A		
6.	Proposed Methodology	A		A		B		A		
7.	System Design	A		A		A		A		
8.	Details of hardware &Software	A		A		A		A		
9.	Implementation details	A		A		A		A		
10.	Result Analysis	B		A		A		A		
11.	Conclusion and Future Work	A		A		A		A		

A: Satisfactory B: Average C: Needs Improvement

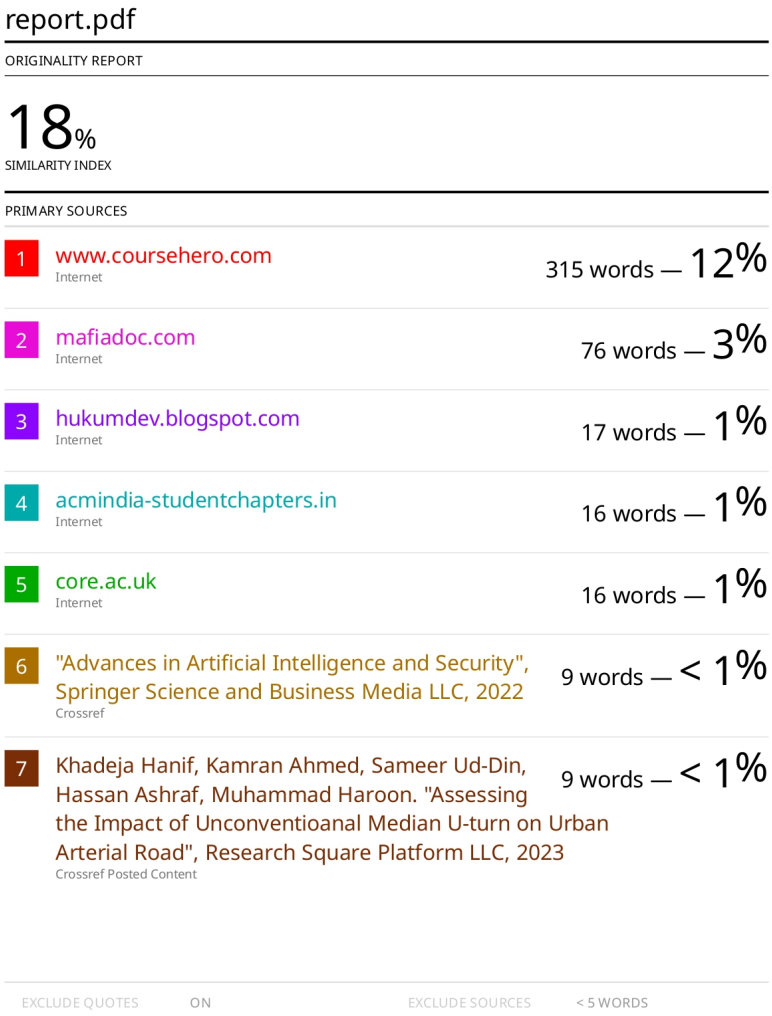
Project Guide Name and Sign

Jyoti P. Vengurlekar

Figure A.1: Weekly Progress Report

Appendix B

Plagiarism Report


report.pdf

ORIGINALITY REPORT

18%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.coursehero.com Internet	315 words — 12%
2	mafiadoc.com Internet	76 words — 3%
3	hukumdev.blogspot.com Internet	17 words — 1%
4	acmindia-studentchapters.in Internet	16 words — 1%
5	core.ac.uk Internet	16 words — 1%
6	"Advances in Artificial Intelligence and Security", Springer Science and Business Media LLC, 2022 Crossref	9 words — < 1%
7	Khadeja Hanif, Kamran Ahmed, Sameer Ud-Din, Hassan Ashraf, Muhammad Haroon. "Assessing the Impact of Unconventioanal Median U-turn on Urban Arterial Road", Research Square Platform LLC, 2023 Crossref Posted Content	9 words — < 1%

EXCLUDE QUOTES ON EXCLUDE SOURCES < 5 WORDS

Figure B.1: Plagiarism Report

Appendix C

Publication Details / Copyright / Project Competitions

1. S. Gaud, S. Kale, R. Sarambale, P. Gunjgur, “A Recommendation System for Integrated Agriculture Using Convolutional Neural Networks with Random Forest Algorithm,” Fifth International Conference on Computational Intelligence and Communication Technologies (CCICT), July 2022. [Published]

Acknowledgments

I thank the many people who have done lots of nice things for me.

Date: _____