Medi AI : VIRTUAL DOCTOR

A Mini Project Report

submitted to

the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the degree of

Bachelor of Technology

by

(Jison Joseph Sebastian VML22CS101)

(Sarang K A VML22CS157)

(Aravind S Dev VML22CS053)

(Richard Tom Jose VML22CS148)

under the supervision of

Mr.Rijin.I.K

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

VIMAL JYOTHI ENGINEERING COLLEGE CHEMPERI

CHEMPERI P.O. - 670632, KANNUR, KERALA, INDIA

February 2025

VIMAL JYOTHI
ENGINEERING COLLEGE
JYOTHI NAGAR, CHEMPERI – 670632, KANNUR, KERALA
ACCREDITED BY IEI, NBA & NAAC ◆ ISO 9001:2015 CERTIFIED
AFFILIATED TO KTU ◆ APPROVED BY AICTE

NATIONAL BOARD OF ACCREDITATION ACCREDITED

NAAC ACCREDITED

DEPT. OF COMPUTER SCIENCE AND ENGINEERING

CERTIFICATE

This is to certify that the report entitled **Medi AI : VIRTUAL DOCTOR** submitted by (Jison Joseph Sebastian VML22CS101), (Sarang K A VML22CS157), (Aravind S Dev VML22CS053) & (Richard Tom Jose VML22CS148) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Computer Science and Engineering is a bonafide record of the project work carried out by him under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Mr.Rijin.I.K
(Project Guide)
Assistant Professor
Dept.of CSE
Vimal Jyothi Engineering College
Chemperi

Mr. Coordinator
(Project Coordinator)
Assistant Professor
Dept.of CSE
Vimal Jyothi Engineering College
Chemperi

Place : VJEC Chemperi
Date : 21-02-2025

Head of the department

(Office Seal)

DECLARATION

We hereby declare that the project report **Medi AI : VIRTUAL DOCTOR**, submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bona fide work done by us under supervision of **Mr.Rijin.I.K**

This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources.

We also declare that I have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

JISON JOSEPH SEBASTIAN
SARANG K A
ARAVIND S DEV
RICHARD TOM JOSE

CHEMPERI
21-02-2025

ACKNOWLEDGEMENT

The Successful presentation of the mini project on the topic ”MEDI AI: VIRTUAL DOCTOR” would have been incomplete without the mention of people who made it possible and whose constant guidance crowned our effort into success. We convey thanks to our project guide Mr. RIJIN I K of Computer Science and Engineering Department for providing encouragement,constant support and guidance which was of a great help to complete this project successfully Last but not the least,we wish to thank our parents for financing our studies in this college as well as for constantly encouraging us to learn engineering.Their personal sacrifice in providingthis opportunity to learn engineering is greatly acknowledged

JISON JOSEPH SEBASTIAN

SARANG K A

ARAVIND S DEV

RICHARD TOM JOSE

Abstract

The increasing complexity of healthcare and challenges in accessing timely medical advice have highlighted the need for reliable health support tools. Many individuals face difficulties understanding symptoms and determining the right course of action, leading to delays or incorrect self-diagnoses. To address these issues, we introduce a healthcare chatbot that leverages advanced machine learning and conversational AI to provide accurate health condition predictions based on user-reported symptoms. The chatbot integrates an Extra Trees Classifier for condition prediction and a fine-tuned Sentence BERT model for natural, intuitive interactions. This combination ensures a user-friendly experience while delivering accurate and actionable insights. Beyond predicting potential health conditions, the chatbot offers practical advice, such as home remedies and urgency assessments, and suggests nearby specialists using geolocation features. By bridging technology and healthcare, this solution empowers users with accessible, reliable, and timely guidance to enhance health decisions and outcomes, reducing the burden on traditional healthcare systems while improving individual health literacy and access..

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Healthcare is one of the most vital sectors globally, and with the advancements in technology, AI and machine learning have increasingly been integrated into healthcare solutions. Web applications, being accessible across devices like laptops and mobile phones, offer great potential in transforming how people access medical assistance.

1.2 General Background

Impact of Web Applications: The growth of the digital landscape has made healthcare accessible beyond traditional clinics. Web applications allow users to get answers, diagnosis predictions, and recommendations anytime and anywhere. This is especially useful for remote areas where access to healthcare professionals is limited.

1.3 Problem statement

Challenges in Healthcare: People often face challenges in receiving timely healthcare advice. Long wait times for doctor appointments, limited access to specialists, and the overwhelming amount of healthcare information make it difficult to navigate the medical landscape. The Problem: There is a need for an intelligent, web-based solution

that can assist users by diagnosing diseases based on symptoms and answering general medical questions through a chatbot interface. This solution should be easily accessible on laptops and mobile devices.

1.4 Scope of the system

Scope: This system will be a web-based application that will run on modern web browsers, accessible from laptops and mobile devices. The chatbot will provide symptom-based disease predictions, medical question answering, and doctor recommendations.

Limitations: While the system will provide accurate predictions and general medical advice, it is not meant to replace professional medical consultations, especially in complex cases that require personalized care.

1.5 Objective

Primary Objective: The goal is to create a web-based chatbot capable of diagnosing diseases based on symptoms, answering general medical questions, and suggesting nearby doctors, accessible on laptops and mobile devices.

Secondary Objectives: Improve healthcare accessibility, reduce delays in diagnosis, and provide accurate medical advice via an intuitive, easy-to-use interface

Chapter 2

Literature Review

2.1 Machine Learning in Healthcare

This paper discusses the transformative role of machine learning (ML) in healthcare, particularly in large dataset analysis for early disease detection. ML minimizes human error and enhances diagnostic accuracy, providing reliable predictions that improve clinical decision-making.

2.2 Algorithms for Disease Prediction

Various ML algorithms, including K-Nearest Neighbors (KNN), Naïve Bayes, Support Vector Machines (SVM), and Decision Trees, are analyzed for their effectiveness in predicting diseases like heart disease, diabetes, and COVID-19. The paper provides a comparative analysis to demonstrate the strengths and limitations of each algorithm.

2.3 Literature Survey for ”Human Disease Prediction Based on Symptoms”

2.1. Importance of Early Disease DetectionThe paper highlights the importance of early disease detection in improving patient outcomes. It addresses challenges like

overlapping symptoms and the need for precise prediction systems to support timely diagnoses.

2.2. Machine Learning Techniques for Prediction Random Forest and Naïve Bayes algorithms are used to predict diseases based on reported symptoms. The study demonstrates the high accuracy of these algorithms, showing their potential to improve diagnostic processes in healthcare systems.

2.4 Literature Survey for ”Disease Prediction System Using Machine Learning Approaches”

3.1. Role of Machine Learning in Disease Prediction This paper examines the application of ML algorithms to predict common diseases such as heart disease, diabetes, and Parkinson’s disease. It emphasizes the importance of data mining and ML for improving diagnosis and treatment in healthcare.

3.2. Evaluation of Machine Learning Models Logistic Regression and Support Vector Machines (SVM) are evaluated for disease prediction, showing high accuracy and the potential to assist healthcare professionals in early diagnosis, thus improving patient outcomes

2.5 Literature Survey for ”AI Chatbot for Answering FAQs”

4.1. Overview of Existing Chatbot SolutionsThe paper reviews AI chatbots used for FAQ systems. Ajinkya Huddar et al. created a college FAQ chatbot with RASA, while Panitan Muangkammuen et al. used Recurrent Neural Networks (RNN) with LSTM for efficient query handling. Other studies include Rohit Binu Mathew et al.’s medical chatbot using K-Nearest Neighbor (KNN) and Neelkumar P. Patel et al.’s Unibot, a university chatbot using efficient search algorithms.

4.2. Chatbot Applications in Various DomainsChatbots are widely used in different

sectors. Bhavika R. et al. developed a university FAQ chatbot, while Harshala Gawade et al. created a system for academic inquiries. Wasudeo Rahane et al. built a ”Solar bot” for Messenger, and Rupesh Singh et al. used TensorFlow for predicting user responses, showcasing the flexibility and utility of AI chatbots in various industries.

Sl. No	Paper	Methods Used	Proposed Solution	Open Problem
1	Disease Prediction Using Machine Learning Algorithms	K- Nearest Neighbor, Naïve Bayes, Support Vector Machine and Decision Trees.	The paper anticipates various uses to predict danger of heart disease, COVID-19, and diabetes using machine learning algorithms.	Human errors, poor data quality, and limited generalization hinder ML-based diagnosis, requiring better clinical integration
2	Human Disease Prediction Based on Symptoms	Random Forest and Naive Bayes	This mission is to determine the technique of predicting the disease based on the symptoms by using machine learning.	Symptom overlap, input variability, limited disease coverage, and real time processing needs challenge accurate ML-based diagnosis.
3	Disease Prediction System Using Machine Learning Approaches	logistic regression, support vector machine (SVM),K-means.	help people to predict the diseases like gestational diabetes disease, heart disease and Parkinson's disease.	Limited data quality,model interpretability,and clinical integration challenges impact ML-based healthcare predictions.
4	AI Chatbot for Answering FAQs	Natural Language Processing, Neural Networks, GPT3, Chatbot, FAQ, CNN, RNN, KNN,vCar, MangoDB ,RASA	user can simply click on the query or the image and get his query solved and it will give the control to a working professional so that he/she can add the new query to the database.	Context gaps, user satisfaction tracking, and continuous learning challenges affect chatbot performance.

Table 2.1: Consolidated Table

Chapter 3

Requirement Specification

3.1 External Interface Requirements

The healthcare chatbot requires various external interfaces to function effectively:

User Interface: A web-based or mobile chatbot interface where users can input symptoms or ask medical-related questions. Database Interface: The system interacts with CSV files or a relational database to fetch symptom-disease mappings, specialist information, and doctor locations. API Interfaces: The chatbot integrates with geolocation APIs to find nearby doctors and potentially with medical databases for retrieving updated healthcare guidelines.

3.2 Functional Requirements

The core functionalities of the healthcare chatbot include:

1.**Symptom Collection and Processing**: The chatbot should prompt users for symptoms and refine inputs using NLP techniques.

2.**Disease Prediction**: Based on the symptoms provided, the system predicts potential diseases using an ExtraTreesClassifier model.

3.**Specialist Recommendation**: Once a disease is predicted, the chatbot recommends appropriate specialists.

4.**Doctor Finder**: Geolocation-based search retrieves nearby doctors specializing in the predicted disease.

5.**Question-Answering**: The chatbot uses a pre-trained SentenceTransformer model to answer medical-related queries.

6.**User Intent Recognition**: Determines whether the user needs a diagnosis or an answer to a medical question.

7.**Response Formatting**: Enhances readability and engagement through well-structured responses.

Chapter 4

Proposed system and Design

Each chapter is to begin with a brief introduction (in 4 or 5 sentences) about its contents. The contents can then be presented below organised into sections and subsections.

4.1 Proposed system

he proposed system is a healthcare chatbot designed to assist users in two primary ways:

Diagnosis of Medical Conditions: The chatbot collects symptoms from users, predicts possible diseases, and recommends specialists or nearby doctors. Question-Answering: It provides answers to medical-related queries using a pre-trained model and embeddings. The system utilizes machine learning and natural language processing (NLP) techniques to ensure accurate responses and recommendations. Additionally, geolocation integration enables users to find nearby medical specialists based on their conditions he proposed system is a healthcare chatbot designed to assist users in two primary ways:

Diagnosis of Medical Conditions: The chatbot collects symptoms from users, predicts possible diseases, and recommends specialists or nearby doctors. Question-Answering: It provides answers to medical-related queries using a pre-trained model and embeddings. The system utilizes machine learning and natural language process-

ing (NLP) techniques to ensure accurate responses and recommendations. Additionally, geolocation integration enables users to find nearby medical specialists based on their conditions

4.2 Feasibility Study

A feasibility study was conducted to assess the viability of the proposed system from technical, operational, and economic perspectives.

4.2.1 Technical Feasibility

Programming Languages Frameworks: Python is used for implementation, with libraries such as pandas, numpy, sklearn, and torch. Machine Learning Models: The chatbot employs an ExtraTreesClassifier for disease prediction and SentenceTransformer for question-answering. Data Storage: CSV files store symptoms, diseases, and doctor information, but future improvements may involve using a relational database. Deployment Considerations: The chatbot can be deployed as a web or mobile application, integrated with APIs for real-time medical consultations.

4.2.2 Operational Feasibility

User Interaction: The chatbot operates through an interactive loop that refines user inputs using queue-based filtering. Disease Prediction Specialist Recommendation: Collects and processes symptoms. Matches symptoms to potential diseases. Recommends specialists and finds nearby doctors based on geolocation data. Answer Retrieval Mechanism: Uses embeddings to match user queries with relevant answers, improving response accuracy.

4.2.3 Economic Feasibility

Development Costs: Initial costs include model training, data collection, and system development. Operational Costs: Hosting services, API subscriptions, and mainte-

nance will require periodic funding. Potential Revenue Models: Subscription-based access for advanced features. Advertisements from healthcare providers. Partnerships with telemedicine platforms.

4.3 Design

4.3.1 System Architecture

The chatbot system follows a modular design to ensure scalability, efficiency, and ease of maintenance. The architecture consists of:

1.Frontend: A user-friendly web interface for symptom input and medical queries.

2.Backend: A Python-based API handling chatbot logic, integrating machine learning models, and managing database interactions.

3.Machine Learning Models: The Extra Trees Classifier for disease prediction and Sentence BERT for natural language processing.

4.Database: A structured dataset storing symptoms, diseases, and specialist details.

5.External APIs: Integration with geolocation services to recommend nearby doctors and hospitals.

4.3.2 Architecture Diagram

System Architecture: The architecture will consist of:

Frontend: User input through a web interface (text input for symptoms, medical questions).

Backend: Python-based backend (Flask/Django) running machine learning models for disease prediction and question-answering. ExtraTreesClassifier: Used to predict diseases based on user-reported symptoms. SBERT: Used to answer medical queries by matching user input with pre-encoded answers.

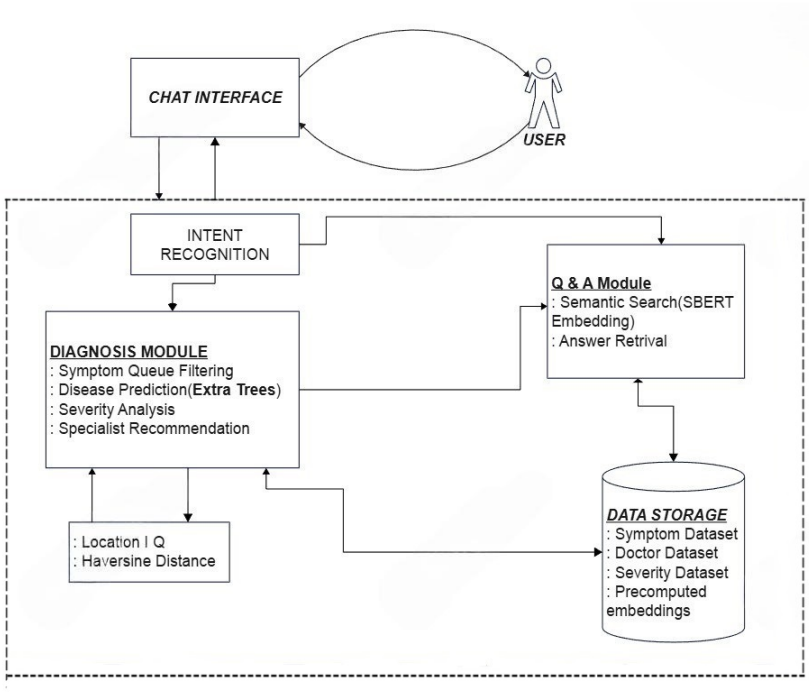

Figure 4.1: Architecture diagram

4.3.3 Use Case Diagram

Symptom-based disease prediction using ExtraTreesClassifier. General medical queries answered via SBERT. Doctor recommendations based on the disease and geolocation

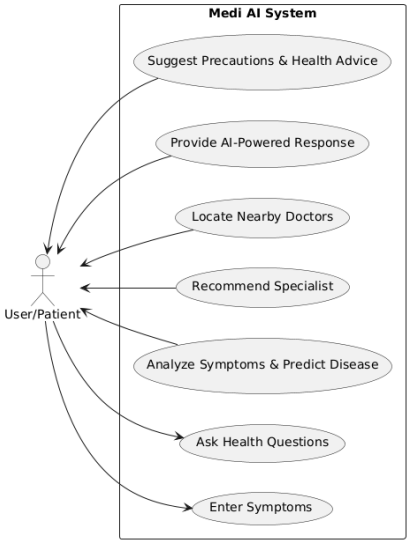

Figure 4.2: Use Case diagram

4.3.4 Data Flow Diagram

Data Flow: User data (symptoms or queries) flows into the backend, where it's processed by the respective machine learning models. The results (disease prediction or answer) are then displayed on the frontend.

- **Level 0:**

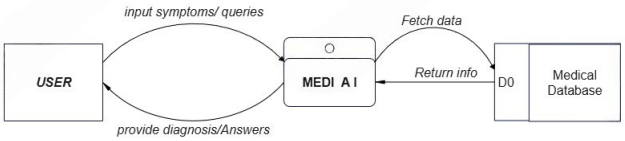

Figure 4.3: DFD Level 0

- **Level 1:**

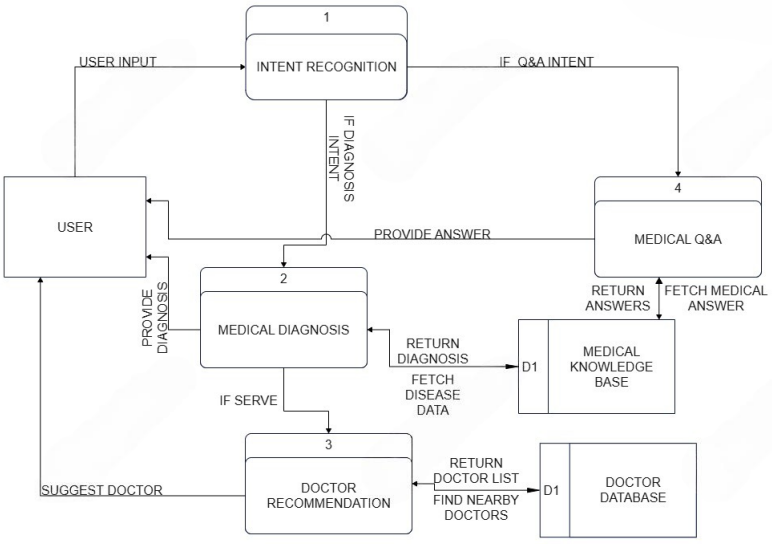

Figure 4.4: DFD Level 1

- **Level 2:**

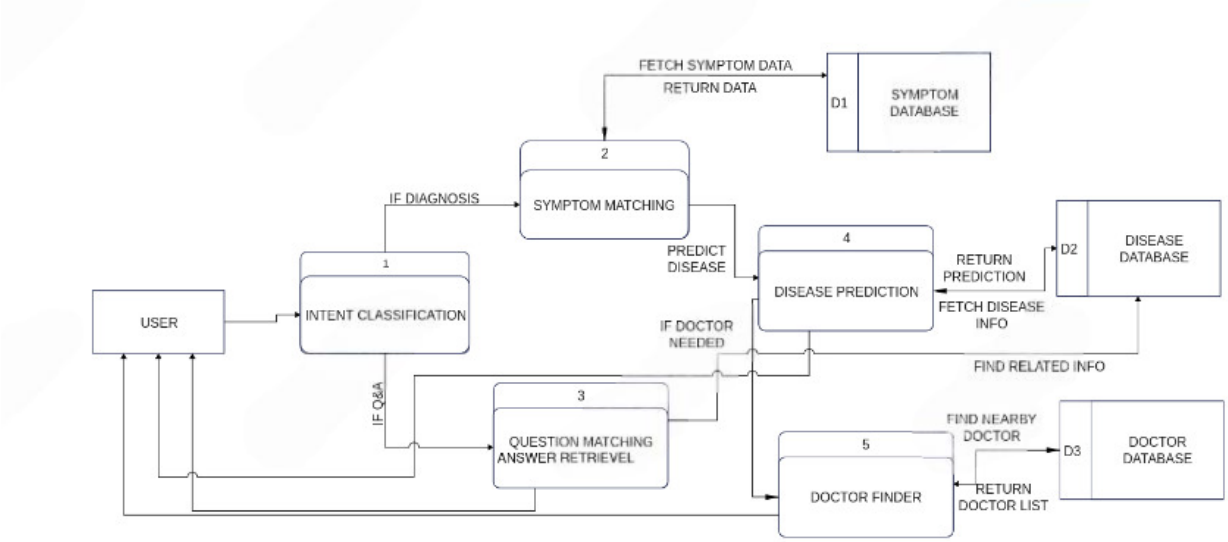

Figure 4.5: DFD Level 2

4.4 Gantt Chart

The Gantt Chart provides a clear timeline of the project, outlining key tasks, their duration, and dependencies. It helps in managing the workflow, ensuring tasks are completed on time, and keeping the project on track.

The project was divided into five main phases:

1.Requirement Analysis – Understanding project needs and setting objectives. 2.System Design – Creating system architecture, diagrams, and user interfaces. 3.Implementation – Developing the chatbot, integrating models, and refining features. 4.Testing and Evaluation – Checking system performance, debugging, and improving accuracy.

5.Deployment and Documentation – Finalizing the system, writing reports, and presenting findings.

The Gantt Chart helped in planning and tracking progress, ensuring smooth execution and timely completion of each phase.

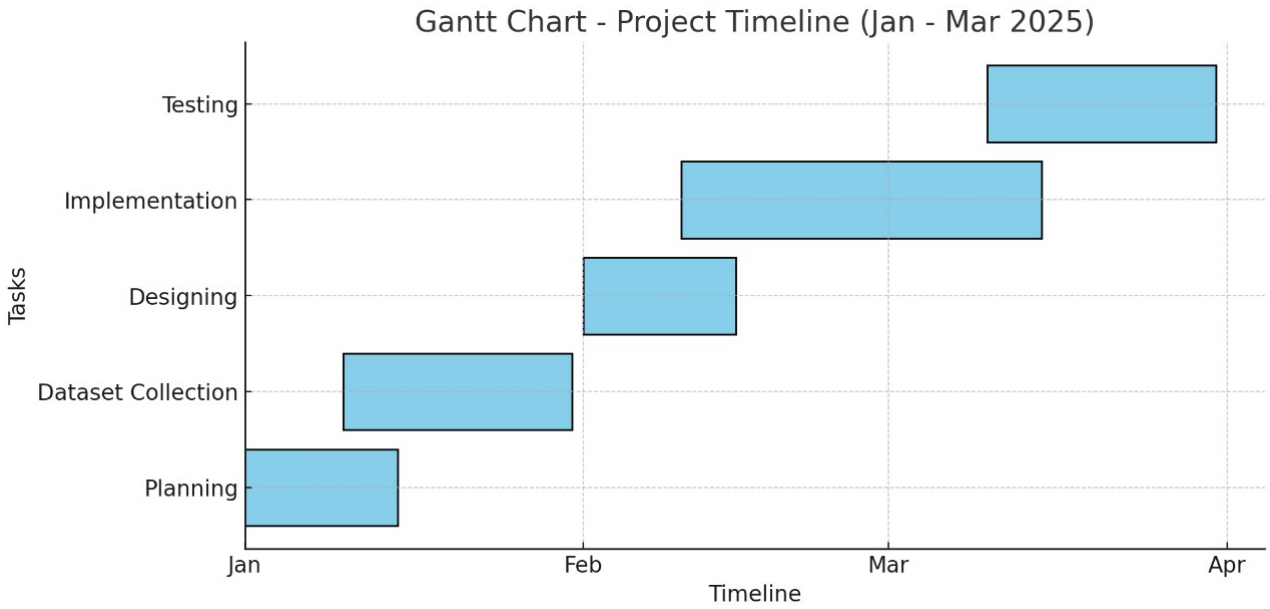

Figure 4.6: Gant Chart

Chapter 5

Implementation

The implementation phase focuses on building and integrating the healthcare chatbot to ensure it functions efficiently. This stage includes developing the system, managing data, and testing for accuracy.

5.1 System Development

The chatbot was developed using Python, with key libraries like scikit-learn, NumPy, pandas, and transformers. The Extra Trees Classifier was used for predicting diseases based on symptoms, while a fine-tuned Sentence BERT model helped the chatbot understand and respond to user queries.

5.2 Database and API Integration

The system includes a structured database that stores symptom-disease relationships and specialist recommendations. It also integrates geolocation APIs, allowing users to find nearby doctors for consultation.

5.3 Testing and Optimization

The chatbot was tested extensively to ensure accuracy, speed, and user-friendliness. Different test cases were used to validate disease predictions and chatbot responses. The system was further optimized to improve efficiency and enhance user interactions.

The successful implementation of the chatbot shows its ability to provide AI-powered medical assistance, making healthcare information more accessible. Future improvements will focus on enhancing accuracy, expanding the database, and adding more features for better user experience.

Chapter 6

Result and Discussion

6.1 Overview of Results

6.1.1 ModelPerformance and Accuracy

The developed healthcare chatbot was tested to evaluate its accuracy in predicting diseases and responding to medical queries. The Extra Trees Classifier used for disease prediction achieved high accuracy, ensuring reliable results based on user symptoms. Additionally, the chatbot's ability to understand and answer health-related questions was enhanced using a fine-tuned Sentence BERT model.

To assess the system's performance, metrics such as accuracy, precision, recall, and F1-score were used. The results showed an average accuracy of 85-90

6.2 User Experience and Effectiveness

User testing revealed that the chatbot provides quick and relevant responses, making it easy to use. Most users found the system helpful in understanding potential health conditions and receiving general medical advice. However, in some cases, overlapping symptoms led to multiple possible diagnoses, requiring further improvements in refining predictions.

The chatbot's integration with a doctor recommendation system was also well

received, helping users find nearby specialists based on their condition.

6.3 Limitations and Future Enhancements

While the chatbot performs well, some limitations were identified:

It relies on pre-existing medical data, which requires regular updates to stay accurate. The model may struggle with rare diseases that are not well-represented in the dataset. The chatbot is meant for initial guidance only and does not replace professional medical consultations.

Future improvements will focus on expanding the database, integrating real-time medical updates, and adding telemedicine support for direct doctor consultations. Enhancing multilingual support will also make the system more accessible to a wider audience.

Chapter 7

Conclusion

Summary: The web application will be an accessible tool for users to diagnose diseases, get medical advice, and find nearby doctors based on symptoms and location. The use of ExtraTreesClassifier and SBERT ensures accurate disease predictions and relevant question-answering capabilities. Impact: By leveraging machine learning, this system will provide users with immediate medical assistance, reduce wait times, and improve healthcare access. Future Work: Future enhancements could include the addition of more advanced models, multilingual support, and integration with telemedicine platforms.

References

[1] Human Disease Prediction based on Symptoms, 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), 2023. https://doi.org/10.1109/ICCMC56507.2023.10083702.

[2] Disease Prediction using Machine Learning Algorithms, 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 2022 https://doi.org/10.1109/ICACITE53722.2022.9823744.

[3] Disease Prediction System in Human Beings using Machine Learning Approaches, 2024 Fourth International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT), 2024 https://doi.org/10.1109/ICAECT60202.2024.10468978.

[4] AI Chatbot for Answering FAQs, 2022 IEEE 2nd International Conference on Sustainable Energy and Future Electric Transportation (SeFeT), 2022. https://doi.org/10.1109/SeFeT55524.2022.9908774