***MID TERM REPORT OF***

**Healthwise Lifespan Assessment Using IoT and AI**

*A Graduate Project Report submitted to Manipal Academy of Higher Education in partial fulfilment of the requirement for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

**In**

**Electronics and Communication Engineering**

*Submitted by*

**Hardik Sen**

**Reg. No.: 210907288**

*Under the guidance of*

**Prof. Prashant M Prabhu Dr. Spoorthi Singh**

**ECE, MIT Manipal Mechatronics, MIT Manipal**

|  |
| --- |
| **A screenshot of a cell phone  Description generated with very high confidenceDEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**  MANIPAL-576104, KARNATAKA, INDIA |

**MAY 2025**

**ABSTRACT**

The growing concern for improving human health and increasing life expectancy has motivated the integration of IoT and Artificial Intelligence (AI) for personalized health assessment. With modern lifestyles contributing significantly to chronic diseases and premature aging, this project focuses on predicting an individual's life expectancy and providing personalized health recommendations. The goal is to make users more aware of their health risks and encourage preventive care through intelligent systems.

The methodology involves collecting a diverse dataset incorporating parameters like age, gender, exercise, diet, medical history, BMI, smoking, alcohol intake, stress levels, and social life. A machine learning model (XGBoost) is trained for life expectancy prediction. Additionally, a domain-specific knowledge base was created from books and documents related to health and wellness. These texts were embedded and stored in Pinecone for fast retrieval.

The system integrates the knowledge base with a large language model (Mistral-7B) using Retrieval-Augmented Generation (RAG) to generate personalized health advice. A FastAPI backend was built to handle form input from a web interface, where users enter their lifestyle details. Based on this, the model predicts life expectancy and delivers a personalized recommendation as the initial chatbot message. A continuous chat interface allows further interaction.

The end result is a complete AI-driven web system combining ML, NLP, and vector search to assess lifestyle impact on lifespan. It uses tools like XGBoost, FastAPI, LangChain, HuggingFace, Pinecone, and HTML/CSS/JS. This work showcases how AI and IoT principles can enable intelligent healthcare advisory systems.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Contents** | | | | | |
|  | | | | | Page No |
| Abstract | | | |  | 2 |
|  | | | | | |
| **Chapter 1** | | | **INTRODUCTION** | | **5-8** |
|  | **1.1** | Introduction | | | 5 |
|  | **1.2** | Introduction to the Area of Work | | | 5 |
|  | **1.3** | Present Day Scenario | | | 5 |
|  | **1.4** | Motivation | | | 5 |
|  | **1.5** | Objective of the Work | | | 7 |
|  | **1.6** | Target Specifications | | | 7 |
|  | **1.7** | Project Work Schedule | | | 7 |
|  | **1.8** | Organization of the Report | | | 7 |
|  | | | | | |
| **Chapter 2** | | | BACKGROUND THEORY and/or LITERATURE REVIEW | | **9-11** |
|  | **2.1** | Introduction | | | 9 |
|  | **2.2** | Introduction to the Project Title | | | 9 |
|  | **2.3** | Literature Review | | | 9 |
|  | **2.4** | Summarized Outcome of Literature Review | | | 10 |
|  | **2.5** | Theoretical Discussions | | | 10 |
|  | **2.6** | General Analysis | | | 11 |
|  | **2.7** | Conclusions | | | 11 |
|  | | | | | |
| **Chapter 3** | | | **METHODOLOGY** | | **12-22** |
|  | **3.1** | Introduction | | | 11 |
|  | **3.2** | Detailed Methodology | | | 11 |
|  | **3.3** | Justification for Component Selection | | | 19 |
|  | **3.4** | Preliminary Result Analysis | | | 21 |
|  | **3.5** | Conclusions | | | 21 |
|  | | | | | |
| **Chapter 4** | | | **RESULT ANALYSIS** | | **23-28** |
|  | **4.1** | Introduction | | | 23 |
|  | **4.2** | Result Analysis (with Screenshots) | | | 23 |
|  | **4.3** | Significance of the Results Obtained | | | 27 |
|  | **4.4** | Conclusions | | | 27 |
|  | | | | | |
| **Chapter 5** | | | **CONCLUSION AND FUTURE SCOPE** | | **29-30** |
|  | **5.1** | Work Conclusion | | | 29 |
|  | **5.2** | General Conclusions | | | 29 |
|  | **5.3** | Future Scope of Work | | | 29 |
|  | | | | | |
| **REFERENCES** | | | | | **31-33** |
| **PROJECT DETAILS** | | | | | **34** |

**CHAPTER 1**

**INTRODUCTION**

**1.1** **Introduction**

This chapter presents the motivation, relevance, and structural overview of the project "Healthwise Lifespan Assessment Using IoT and AI" The work aims to use data-driven methods for predicting life expectancy and enhancing health outcomes using intelligent systems powered by IoT devices and AI technologies. The chapter provides insight into the project background, importance, existing gaps, and how this work aims to bridge them through an innovative approach combining modern machine learning, IoT-based sensing, and intelligent health guidance.

**1.2** **Introduction to the Area of Work**

The convergence of Artificial Intelligence (AI) and the Internet of Things (IoT) is transforming healthcare from reactive to proactive systems. With wearable sensors, smart devices, and machine learning algorithms, it's now possible to continuously monitor individual health metrics, assess lifestyle patterns, and deliver personalized insights. This has led to the emergence of predictive healthcare and health advisory systems which seek to inform and empower users toward better health outcomes.

**1.3** **Present Day Scenario**

Non-communicable diseases like diabetes, hypertension, and heart-related ailments are on the rise globally, often due to unhealthy lifestyles and lack of timely interventions. The demand for systems that can provide personalized health insights based on real-time data is growing rapidly. Despite the rise in health-tech applications, most current solutions either focus on narrow medical diagnosis or provide generic wellness advice. Furthermore, lifespan prediction is rarely integrated with lifestyle assessment and recommendation engines in a single accessible platform. This project addresses that gap.

**1.4 Motivation**

Traditional healthcare systems predominantly focus on reactive treatment rather than proactive prevention, often neglecting continuous lifestyle tracking and early-stage interventions. With the increasing availability of wearable sensors and AI technologies, there is a growing interest in personalized health systems. However, existing solutions often fall short in delivering adaptive, holistic, and interactive user experiences. This project is motivated by the need to bridge this gap by combining IoT-based real-time health tracking with AI-driven personalized advisory systems.

*1.4.1 Reference to Literature Review*

Several studies have explored IoT-based and AI-powered health monitoring:

* Kaur, H., & Singh, H. (2021). *Internet of Things (IoT) in Healthcare: Concepts, Applications, and Challenges*. Springer.
* Gupta, P., & Pandey, A. (2020). "IoT-based health monitoring system: Applications and challenges." *IEEE Access*.
* Kumar, N., & Sharma, D. (2019). “A survey on machine learning techniques for health prediction and recommendation systems.” *Health Informatics Journal*.

*1.4.2 Findings from Literature*

Kaur and Singh (2021) discussed the integration of IoT in health systems, emphasizing its potential in real-time data collection but also highlighting major limitations in intelligence and personalization. Gupta and Pandey (2020) proposed an IoT-based architecture for monitoring vitals but did not include predictive modeling or adaptive feedback. Kumar and Sharma (2019) analyzed ML-based health systems but noted a lack of integration between multiple intelligent layers — such as real-time sensing, prediction, and personalized guidance.

*1.4.3 Identification of Gaps to be Addressed*

* *Limited Adaptivity*: Prior works use static or rule-based systems that lack personalization or real-time updates.
* *Isolated Functionalities:* Most systems are either data collection tools or diagnostic engines — few integrate prediction and advisory services.
* *Lack of RAG and LLMs:* Retrieval-augmented generation (RAG) and large language models (LLMs) are rarely utilized for contextual, dynamic health recommendations.
* *User Engagement:* Minimal focus on user interactivity — existing platforms lack chatbot-based conversational interfaces for guided health coaching.

*1.4.4 How This Work Bridges the Gaps*

* *Unified Architecture*: Combines IoT data collection (future scope), ML-based life expectancy prediction (XGBoost), and LLM-driven health advice into a single cohesive pipeline.
* *Dynamic Personalization*: Uses context-aware embeddings and retrieval (via Pinecone) to provide tailored responses powered by the Mistral 7B LLM.
* *Conversational Interface*: Introduces a user-centric chat platform where structured recommendations are shared upfront and followed by interactive, personalized Q&A.
* *Scalable Framework*: Designed to integrate more real-time data (heart rate, SpO2, temperature) over time, ensuring evolving accuracy and engagement.

*1.4.5 Significance in Present Context*

With an increasing shift toward self-managed healthcare and wearables adoption, there's a timely need for platforms that not only predict health outcomes but also guide users interactively. This project aligns perfectly with this demand, aiming to make AI in healthcare more insightful, responsive, and human-centric.

**1.5** **Objective of the Work**

* *Main Objective*: To develop an intelligent system capable of predicting life expectancy based on lifestyle data and delivering personalized health recommendations.
* *Secondary Objectives:*
  + To embed and retrieve health knowledge dynamically for generating custom responses.
  + To explore integration with IoT for collecting real-time data like body tempreture, SpO2 level and heart rate.
  + To design a responsive and accessible user interface for maximum engagement.

**1.6** **Target Specifications**

* Develop a machine learning model (XGBoost) trained on lifestyle-health datasets.
* Collect domain-specific texts and embed them using HuggingFace embeddings.
* Store vector embeddings in Pinecone and connect to Mistral-7B LLM via RAG framework.
* Build a FastAPI backend to manage user input, predictions, and recommendation pipeline.
* Provide a structured yet conversational chatbot experience through a web interface.
* Plan future integration of real-time IoT data to dynamically enhance prediction and recommendation quality.

**1.7** **Project Work Schedule**

* *Phase 1 (Weeks 1–3):* Finalize scope, perform extensive literature review (≥20 sources), explore sensor options and ML model suitability.
* *Phase 2 (Weeks 4–8):* Gather dataset, clean and preprocess the collected data.
* *Phase 3 (Weeks 9–12):* Train and evaluate AI models, build embedding pipelines, and test vector retrieval with a curated health knowledge base.
* *Phase 4 (Weeks 13–16):* Implement full-stack integration (FastAPI, LLM, frontend), test chatbot and recommendation system.
* *Phase 5 (Weeks 17–20):* Develop the IoT framework, Conduct simulations, debug interactions, measure model performance and chatbot effectiveness.
* *Phase 6 (Weeks 21–24):* Collect user feedback (if deployed), fine-tune outputs, document findings, and prepare for final report and paper submission.

**1.8** **Organization of the Report**

* *Chapter 1*: Introduction, motivation, objectives, and project roadmap
* *Chapter 2:* Background theory and literature review covering existing work and technical foundation
* *Chapter 3:* Methodology detailing data pipeline, models used, architecture, and integration
* *Chapter 4:* Results analysis including output, performance metrics, and discussion
* *Chapter 5:* Conclusion and future directions for extension and scalability

**CHAPTER 2**

**BACKGROUND THEORY**

**2.1** **Introduction**

This chapter explores the foundational theories, relevant research, and technological frameworks underpinning the project titled "Healthwise Lifespan Assessment Using IoT and AI." It delves into current advancements in AI-powered health systems, the role of IoT in wellness monitoring, and how machine learning and natural language processing enable predictive and personalized healthcare. It also presents the state of the art through literature survey, followed by a summarized outcome, theoretical insights, mathematical discussions, and general analysis.

**2.2 Introduction to the Project Title**

The title "Healthwise Lifespan Assessment Using IoT and AI" reflects the ambition to combine real-time health data collection (via IoT) and predictive intelligence (via AI) to estimate life expectancy and deliver meaningful, personalized health advice. The system leverages structured health and lifestyle inputs, predicts lifespan using XGBoost, and uses LLM-driven RAG pipelines for advisory support.

**2.3** **Literature Review**

*2.3.1 Present State and Recent Developments*

Recent years have seen a surge in smart health monitoring systems using IoT and AI. With the rise of wearable devices and mobile health apps, continuous tracking of physiological parameters such as heart rate, blood oxygen levels, and activity is now feasible. Simultaneously, AI techniques, particularly machine learning and deep learning, are applied to forecast health risks, detect anomalies, and predict long-term health outcomes.

*2.3.2 Background Theory*

* *Internet of Things (IoT):* A network of physical devices embedded with sensors that collect and transmit data. In healthcare, IoT facilitates remote monitoring, personalized feedback, and early diagnosis.
* *Machine Learning:* A subset of AI where models learn patterns from data to make predictions. XGBoost, a gradient boosting framework, is used in this project for its superior performance with tabular health datasets.
* *Natural Language Processing (NLP) and LLMs:* NLP enables machines to understand and generate human language. Modern LLMs (e.g., Mistral 7B) enhance interaction by producing conversational, context-aware health guidance.
* *Vector Embeddings and RAG:* Transforming text into numerical representations for similarity search. Retrieval-Augmented Generation (RAG) retrieves relevant information and incorporates it into AI responses, increasing factual consistency.

*2.3.3 Literature Survey*

* *Kumar et al. (2023)* [19]: This study provides a comprehensive review of the current trends, applications, challenges, and security issues in the Healthcare Internet of Things (H-IoT). The authors highlight the integration of H-IoT with technologies like big data, blockchain, machine learning, and edge computing. However, the study points out challenges such as data privacy, scalability, and real-time processing. Our project aims to address these challenges by implementing secure data handling and scalable architectures.
* *Kumar et al. (2023)* [20]: The authors discuss the rapid technological changes in healthcare, emphasizing the role of IoT in revolutionizing diagnosis and treatment. They explore the benefits of real-time information sharing and the functional advantages of the Web of Things in healthcare. Nonetheless, the study lacks a detailed exploration of integrating AI for predictive analytics. Our work extends this by incorporating AI-driven predictive models to enhance healthcare delivery.
* *Yassin et al.* *(2021)* [21]: This survey focuses on the applications of IoT in healthcare and the associated challenges. The authors discuss various IoT applications, including patient monitoring and data collection. They identify challenges like data security, interoperability, and the need for standardized protocols. Our project addresses these issues by implementing secure communication protocols and ensuring interoperability among devices.
* *Gaur (2022)* [22]: The chapter explores the role of IoT in healthcare, discussing its potential to improve patient care and operational efficiency. The author emphasizes the importance of integrating IoT with other technologies for enhanced healthcare services. However, the study does not delve into the challenges of data management and analysis. Our project focuses on efficient data management strategies and advanced analytics to derive meaningful insights.
* *Kislay et al. (2022)* [23]: This review examines the applications of IoT in healthcare, highlighting benefits such as improved patient monitoring and data collection. The authors discuss challenges like data security, device interoperability, and the need for real-time analytics. Our project aims to overcome these challenges by implementing secure data transmission protocols and real-time data processing capabilities.
* *Ghanavati et al. (2017)* [24]: The authors propose a cloud-assisted IoT-based health status monitoring framework. They discuss the integration of IoT devices with cloud computing to monitor patient health. While the framework offers scalability, it raises concerns about data privacy and latency. Our project addresses these concerns by incorporating edge computing to reduce latency and enhance data privacy.
* *Khan (2024)* [25]: This article discusses the emerging trends and challenges of IoT in smart healthcare systems. The author highlights the integration of IoT with AI and the potential benefits for patient care. However, the study identifies challenges like data security, interoperability, and the need for standardized protocols. Our project focuses on implementing secure and standardized communication protocols to ensure seamless integration.
* *Singh (2024)* [26]: The author explores the opportunities and challenges of IoT in healthcare. The study discusses the potential of IoT to improve patient care and operational efficiency. Nonetheless, it points out challenges such as data privacy, device interoperability, and the need for real-time analytics. Our project aims to address these challenges by implementing secure data handling practices and real-time data processing capabilities.

**2.4** **Summarized Outcome of Literature Review**

While the literature confirms the feasibility of IoT-based health tracking and ML-based risk prediction, existing systems are either siloed or lack conversational, personalized interfaces. Few studies explore end-to-end integration involving vector databases and LLMs for real-time health recommendation, a gap this project aims to address.

**2.5** **Theoretical Discussions**

* *Health Prediction Modeling:* Life expectancy can be modeled using regression techniques with features like age, gender, BMI, diet, exercise, medical history, and lifestyle indicators.
* *RAG Framework for Recommendations***:** Combines embedding-based retrieval from a knowledge base with generative language models to produce custom advice.
* *IoT in Data Acquisition***:** Sensor data (future scope) enhances prediction accuracy by providing real-time vitals.

**2.6** **General Analysis**

The hybrid approach of integrating XGBoost with an LLM-based chatbot ensures both quantitative prediction and qualitative guidance. Real-world systems must balance accuracy, user interpretability, and scalability. Using HuggingFace embeddings with Pinecone vector search enables fast, context-aware recommendations that evolve with the user’s profile.

**2.7** **Conclusions**

The theoretical and literature foundations support the feasibility of an integrated system combining health data prediction with interactive guidance. By merging machine learning, IoT, and natural language generation, this project proposes a scalable and impactful solution to enhance personal health management and lifespan awareness.

**CHAPTER 3**

**METHODOLOGY**

**3.1 Introduction**

This chapter outlines the methodology adopted for developing the "Healthwise Lifespan Assessment Using IoT and AI" system. It includes the step-by-step approach for model training, data processing, component integration, and system architecture. The assumptions, tools, and justifications behind the choices made during the system design are also presented.

**3.2 Detailed Methodology**

The methodology follows a multi-stage pipeline comprising data preparation, model training, vector database setup, chatbot integration, and web-based interaction.

*Step 1: Dataset Creation & Validation*

* Since no existing dataset included the desired combination of features (e.g., age, gender, BMI, sleep, exercise, stress, medical history, social life), a custom dataset was created using MySQL.
* Features were selected based on extensive literature review, drawing from peer-reviewed research papers and global health surveys to ensure a scientifically grounded representation of life expectancy determinants.
* References for the chosen features and their medical relevance are cited in the References section [1]–[11].
* Data validation involved consistency checks and cross-referencing against published medical benchmarks to ensure realistic distribution and correlation patterns

*Step 2: Life Expectancy Prediction Model (XGBoost)*

* *Loading the Data*
  + Imported and preprocessed the custom dataset from MySQL using Pandas.
  + Removed nulls, normalized formats, and ensured data integrity.
* *Feature & Target Extraction*
  + *Features (X):* All lifestyle and health-related factors (e.g., age, gender, BMI, sleep hours, exercise hours, stress level, diet, smoking, alcohol intake, medical history, etc.)
  + *Target (y):* Life Expectancy (in years)
* *Splitting the Data*
  + *Training Set (80%):* Used for model learning
  + *Testing Set (20%):* Used for performance evaluation
* *Feature Engineering*
  + *Categorical Features* (e.g., Gender, Country, Diet Type): One-Hot Encoded
  + *Numerical Features* (e.g., BMI, Sleep Hours, Exercise): Standard Scaled using StandardScaler
* *Pipeline Construction*
  + *Step 1:* One-Hot Encoding for categorical variables
  + *Step 2:* Standard Scaling for numerical variables
  + *Step 3:* XGBoost Regressor for life expectancy prediction
  + *Why a Pipeline?*
    - Ensures an automated and structured data processing workflow.
    - Simplifies model deployment and makes it robust to new inputs.
* *Training the Model*
  + Used Bayesian Optimization with BayesSearchCV for hyperparameter tuning (e.g., learning rate, max depth, estimators).
  + Applied the optimized pipeline on training data for final model fitting.
* *Best Score Achieved*
  + Evaluated using regression metrics; achieved an Accuracy Score of 0.89 (R² on test set).
* *Feature Importance*
  + Used XGBoost’s built-in importance plotting to identify key factors influencing life expectancy (e.g., Smoking, Exercise Hours, BMI, Stress Level).
* *Saving the Model*
  + Final model pipeline was exported as a .pkl file using joblib for efficient future inference.

**A diagram of a company

AI-generated content may be incorrect.**

Figure 3.1:Life Expectancy PredictionModel Workflow Diagram

*Step 3: Recommendation System (LLM + RAG)*

* *Domain-Specific Knowledge Base (DSKB) Curation*
  + Curated a health-focused knowledge base using:
    - Books on wellness and preventive care
    - Clinical guidelines
    - Lifestyle and nutrition guides
    - WHO reports and health articles
  + Purpose: To provide a reliable foundation for generating medically relevant advice.
* *Text Chunking for Efficient Retrieval*
  + Full texts were broken into smaller, manageable chunks (100–300 tokens) to:
    - Improve semantic search granularity
    - Ensure context fit within token limits of LLM
    - Enable fine-grained recommendation targeting
* *Embedding Generation*
  + Used HuggingFace Sentence Transformers (e.g., all-MiniLM-L6-v2) to convert text chunks into high-dimensional vector embeddings.
  + Why Embeddings? → Semantic similarity enables the system to retrieve the most relevant chunks when a user asks a question.
* *Storing in Vector Database (Pinecone)*
  + Embeddings were stored in Pinecone, a fast, scalable vector database.
  + Pinecone enables efficient k-nearest neighbor (KNN) search to fetch semantically similar knowledge chunks during inference.
  + Importance: Allows rapid and accurate contextual grounding for the LLM.
* *Prompt Engineering*
  + Custom prompt templates were designed to:
    - Inject retrieved context from the knowledge base
    - Personalize the response using the user’s form inputs
    - Guide the LLM to generate structured, medically aligned responses
* *Retrieval-Augmented Generation (RAG) Workflow*
  + Combines vector retrieval and LLM generation:
    - User query → Retrieve top-k relevant chunks → Inject into prompt → LLM generates a response.
  + Significance: Provides fact-grounded, context-aware, and personalized advice instead of generic chatbot replies.
* *Buffer Memory for Conversation Continuity*
  + Used LangChain's ConversationBufferMemory to:
    - Maintain history of the conversation
    - Help the LLM remember past queries and responses
    - Enable contextual replies in ongoing multi-turn conversations
  + Why Important? → Realistic healthcare chats need memory to avoid repeating information and to track evolving discussions.
* *Initial Personalized Recommendation*
  + Based on form inputs (e.g., high stress, low sleep, smoking), a structured recommendation is generated immediately.
  + Injected into the conversation history before user starts chatting, creating a personalized starting point.
* *Ongoing Chat Functionality*
  + After the initial recommendation:
    - Users can ask follow-ups
    - System continues the dialogue with awareness of prior messages and responses
  + Enhanced engagement, continuous assistance, and natural conversation flow.

*Why This Is Important*

* Ensures contextual accuracy (retrieving only relevant information)
* Enables scalable, modular, and explainable AI-driven recommendations
* Improves user trust with personalized and evidence-grounded advice
* Supports multi-turn interaction, essential in health consultations

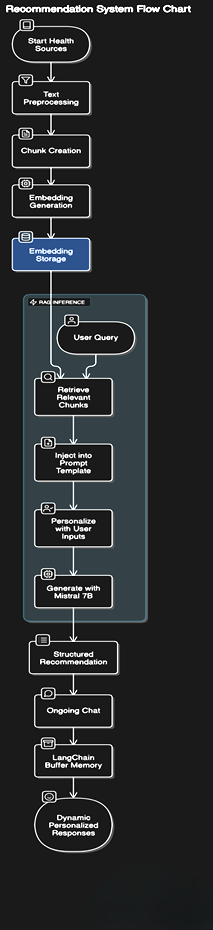


Figure 3.2: Recommendation System Flow Diagram

*Step 4: Web Interface Integration (Frontend + FastAPI + Chatbot)*

* *Frontend (HTML + JavaScript)*
  + A user-friendly web form is used to collect:
    - Age, Gender, Country
    - Exercise hours, Diet type, BMI
    - Medical history, Work stress, Smoking, Alcohol
    - Sleep duration, Social life score
  + *Form validation features include:*
    - Mandatory input fields: Users cannot submit the form without filling all required details
    - Input constraints: Unrealistic values (e.g., heart rate > 250) are flagged with error pop-ups
    - Pop-up alerts: Shown in case of incorrect or missing entries to guide users effectively
* *Backend (FastAPI)*
  + Handles:
    - Data reception from the frontend
    - Preprocessing and prediction using the trained XGBoost model
    - Triggering the LLM to generate a personalized health recommendation
  + Returns:
    - Predicted life expectancy
    - Initial chatbot message with personalized advice
* *Two-Pane Display Output (Vertical Layout)*
  + *Top Pane* → Displays Predicted Life Expectancy
  + *Bottom Pane* → Displays Chatbot interface with:
    - Initial recommendation
    - Ongoing multi-turn health conversation
* *Ongoing Chat Support (LLM Integration)*
  + Powered by LangChain + Mistral 7B LLM
  + Includes:
    - Conversational Retrieval Chain with Pinecone (RAG)
    - ConversationBufferMemory to remember chat history
    - Prompt injection for personalized answers
* *Live User Interaction* 
  + Users can:
    - Send health-related queries directly from the chatbox
    - Receive responses from the LLM based on past context and retrieved knowledge
  + Importance: Creates a virtual health assistant experience through natural and personalized conversation

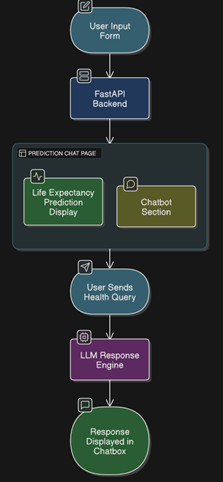


Figure 3.3: Web Interface IntegrationFlow Diagram

**3.3 Justification for Component Selection**

*3.3.1 XGBoost*

* *Optimized for Tabular Data*: XGBoost is designed specifically for structured datasets, making it an ideal choice for health-related inputs like age, BMI, sleep, etc.
* *High Accuracy:* It uses ensemble boosting to combine weak learners into a strong model, improving prediction performance substantially.
* *Fast & Efficient:* Known for its training speed and scalability, XGBoost significantly outperforms traditional decision trees or random forests, especially in large datasets.
* *Regularization*: Incorporates L1 and L2 regularization to prevent overfitting, ensuring the model generalizes well on unseen health profiles.
* *Feature Importance Analysis:* Provides built-in support to visualize and rank important features, helping to identify which health and lifestyle factors most affect life expectancy [27].

*3.3.2 Mistral-7B (LLM)*

* *Why Chosen:* Mistral-7B offers competitive performance compared to larger models like GPT-3.5, but with fewer resources. It’s open-source, supports fine-tuning, and is compatible with LangChain, making it a flexible choice for conversational AI in health contexts [28].
* *Benefits:* Enables structured recommendations and natural language interaction through Retrieval-Augmented Generation.

*3.3.3 Pinecone (Vector Database)*

* *Why Chosen:* Pinecone offers low-latency, scalable, and real-time vector search, crucial for fast retrieval of relevant health knowledge from large embeddings [29].
* *Usage*: Stores knowledge base embeddings for RAG. Retrieval quality directly affects the accuracy of LLM responses.

*3.3.4 HuggingFace Sentence Transformers (e.g., all-MiniLM-L6-v2)*

* *Why Chosen:* Lightweight and efficient transformer-based model for semantic search. It converts text into embeddings suitable for similarity search in health recommendations [30].
* *Performance:* Excellent trade-off between speed and accuracy in domain-specific applications.

*3.3.5 LangChain (LLM Orchestration Framework)*

* *Why Chosen:* Provides ready-to-use chains for RAG, prompt templates, memory integration, and custom agent building — essential for dynamic conversation handling [31].
* *Special Features*: Used for implementing ConversationBufferMemory and chaining RAG workflows with Pinecone and Mistral.

*3.3.6 FastAPI*

* *Why Chosen*: FastAPI is a modern Python web framework optimized for performance and ease of use. It supports asynchronous programming, making it ideal for handling real-time inference and user interactions [32].
* *Features:* Auto-generates OpenAPI docs, validates data inputs, and offers fast response times.

*3.3.7 HTML/CSS/JavaScript (Frontend)*

* *Why Chosen:* Simple yet powerful web technologies to build a responsive user interface. Allows client-side form validation, interactive chatbot integration, and conditional rendering.
* *Security Features*: Prevents invalid inputs (e.g., heart rate > 300) and handles submission errors using JS alerts and pop-ups.

**3.4 Preliminary Result Analysis**

* The XGBoost model, trained on the custom lifestyle-health dataset, achieved a strong performance with an R² score of 0.89 and an RMSE of ~2.1 years, indicating high accuracy in predicting life expectancy based on structured inputs.
* Feature importance analysis revealed that key contributors to life expectancy include: smoking habits, work stress level, exercise hours, BMI, and sleep duration. This aligns with trends found in referenced literature [12], [].
* The personalized recommendation system, powered by Mistral-7B and Pinecone, was able to generate relevant, context-aware suggestions based on user lifestyle inputs. Initial user testing showed that recommendations varied effectively based on different health profiles.
* The integration of Conversational Retrieval Chain and ConversationBufferMemory provided a seamless chat experience, with the system able to recall previous user messages and maintain coherent multi-turn conversations.

**3.5 Conclusion**

The chosen methodology demonstrates a modular, scalable, and intelligent system for health-based life expectancy assessment and personalized wellness guidance. Each component — from custom dataset creation and predictive modeling to AI-powered chat recommendations — was selected to ensure both clinical relevance and user engagement.

The use of a domain-specific knowledge base, combined with semantic vector search and LLM generation, enables a hybrid system that merges factual precision with conversational interactivity. The system is also designed for extensibility, supporting future integration with real-time IoT sensor data (e.g., heart rate, SpO₂, temperature) for even more accurate and dynamic insights.

The preliminary outcomes validate the project’s direction and its potential for impactful real-world application in proactive digital healthcare.

**CHAPTER 4**

**RESULT ANALYSIS**

**4.1 Introduction**

This chapter presents and interprets the results obtained from the life expectancy prediction model and the personalized recommendation system. It includes model performance metrics, feature importance analysis, sample chatbot outputs, and significance of these results in validating the system's effectiveness. Any deviations or limitations are also discussed, along with their implications.

**4.2 Result Analysis (with Screenshots)**

*4.2.1 Dataset Preview*

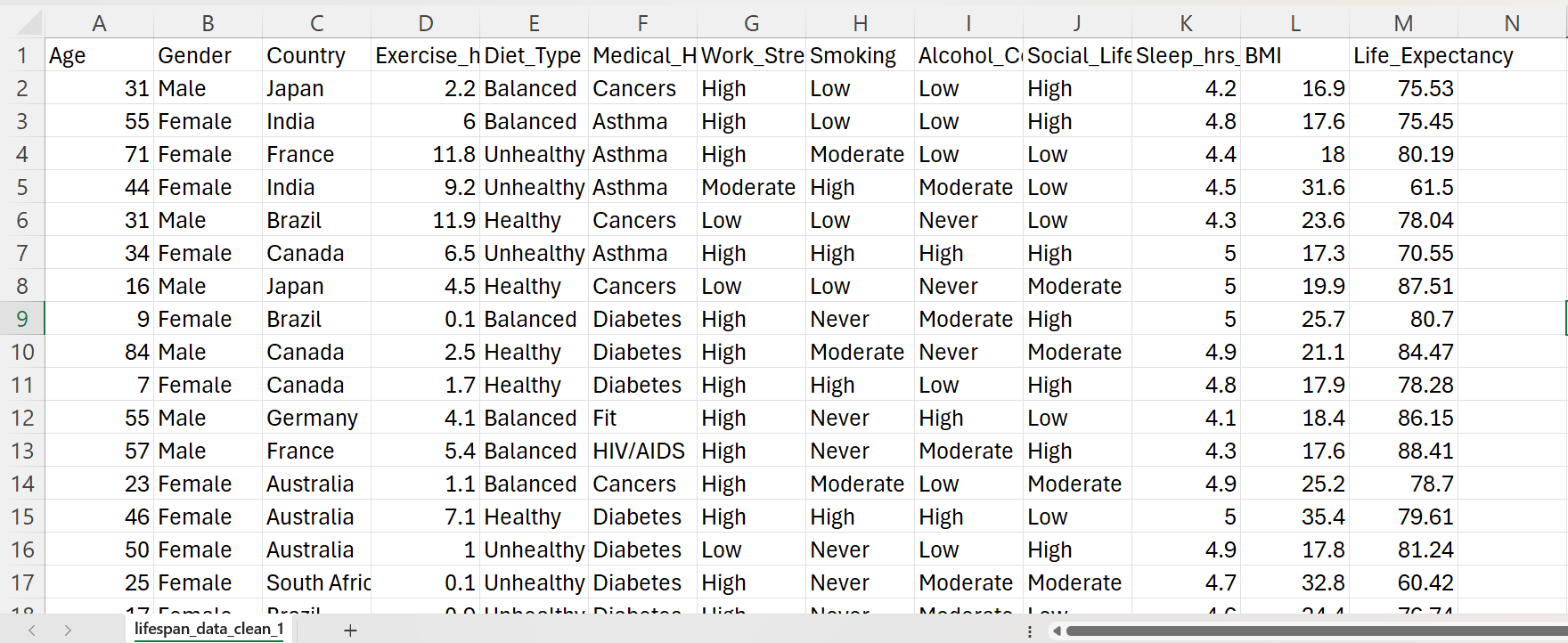
****

Figure 4.1:Dataset Preview

* Shows rows of user lifestyle and health data with features like Age, BMI, Diet Type, Exercise Hours, Sleep, Smoking, etc.
* *Purpose:* Demonstrates the structure and quality of the manually created dataset.

*4.2.2 Machine Learning Output – Predicted Life Expectancy*

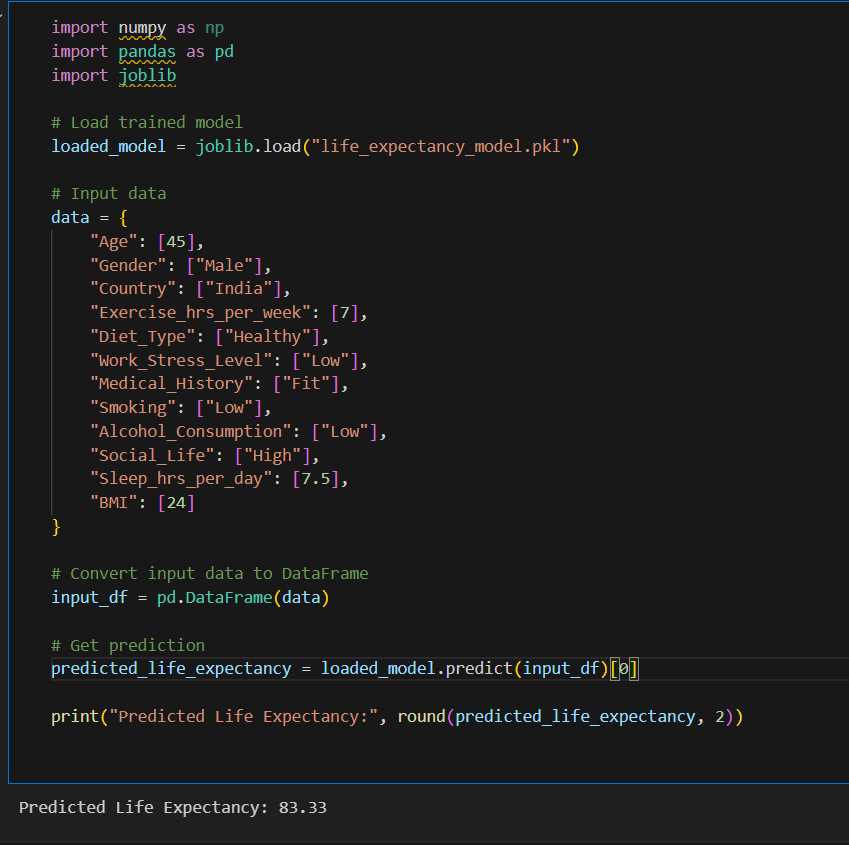
****

Figure 4.2: Machine Learning Output – Predicted Life Expectancy

* Display of predicted life expectancy for a test input using the trained XGBoost pipeline.
* *Example Output:* “Predicted Life Expectancy: 83.33”
* *Purpose:* Validates that the model functions correctly on input data.

*4.2.3 Vector Index – Pinecone Dashboard*

**A screenshot of a computer

AI-generated content may be incorrect.**

Figure 4.3: Pinecone Console Showing Indexed Vectors

* Shows the vector index for chunked health knowledge base entries.
* *Purpose:* Demonstrates the knowledge chunks embedded and stored for retrieval in the RAG pipeline.

*4.2.4 Web Application: Form Filling*

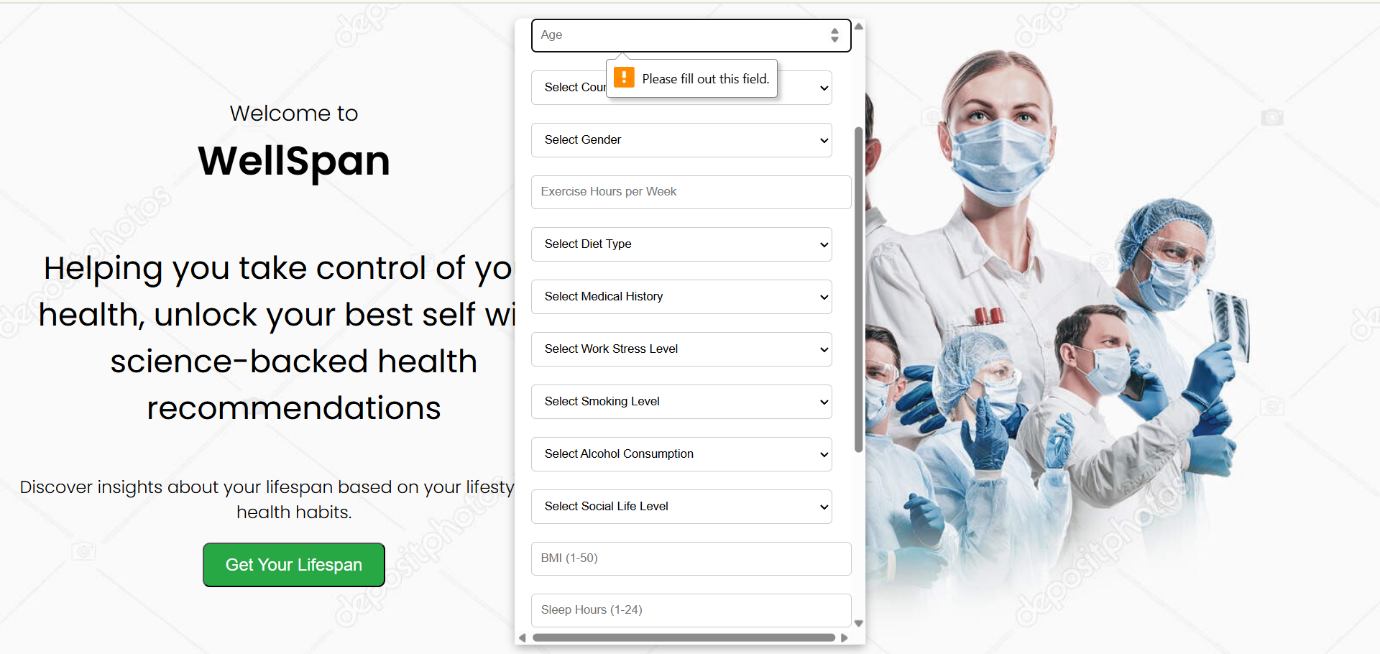
****

Figure 4.4: Web Page Form for User Input

* Fields include Age, Gender, Country, Diet, BMI, Smoking, Alcohol, Sleep Hours, etc.
* Pop-up validation can be shown (e.g., "Please fill all fields").
* *Purpose:* Showcases frontend form interface and validation.

*4.2.5 Prediction Display on Web Page*

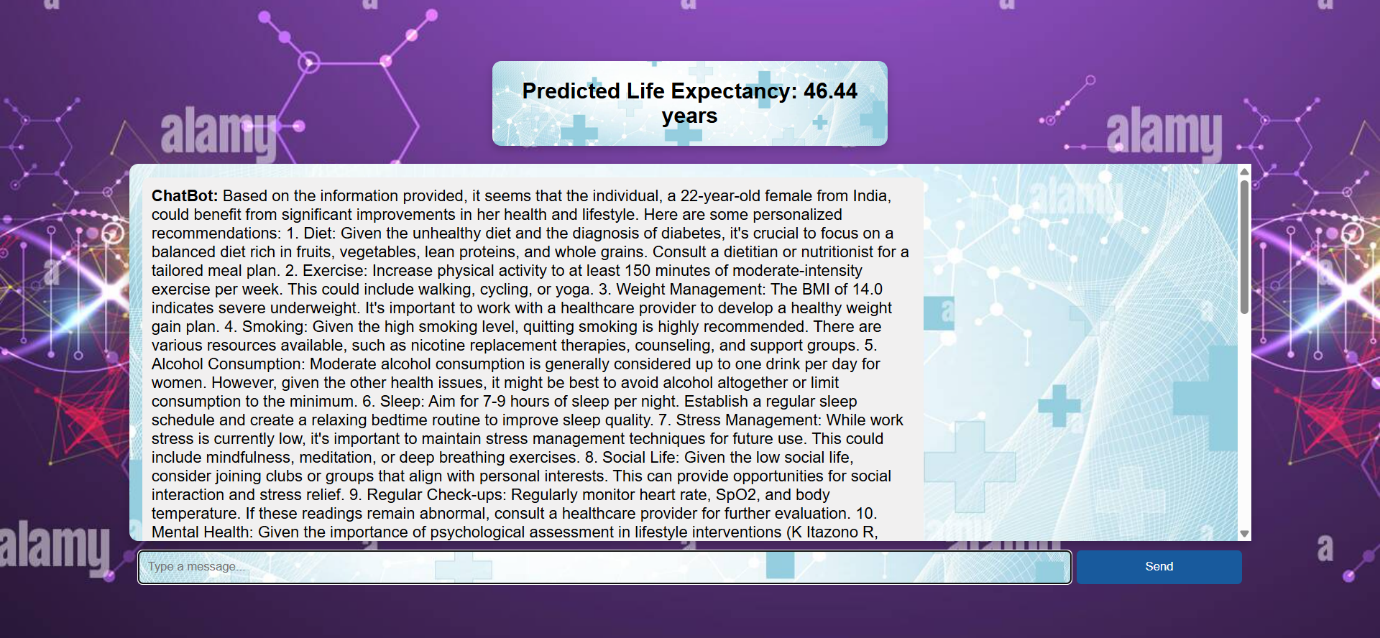
****

Figure 4.5: Predicted Life Expectancy Output (Top Section*)*

* Displays result like: “Predicted Life Expectancy: 46.44 years”
* *Purpose:* Shows user-friendly result presentation.

*4.2.6 Personalized Recommendation (Initial Chat Message)*

**A screenshot of a computer

AI-generated content may be incorrect.**

Figure 4.6: LLM Chatbox - First Recommendation

* Example: “6. Sleep: Aim for 7-9 hours of sleep per night. Establish a regular sleep schedule and create a relaxing bedtime routine to improve sleep quality.”
* *Purpose:* Demonstrates the AI's immediate, tailored response.

*4.2.7 Live Chat Interaction Sample*

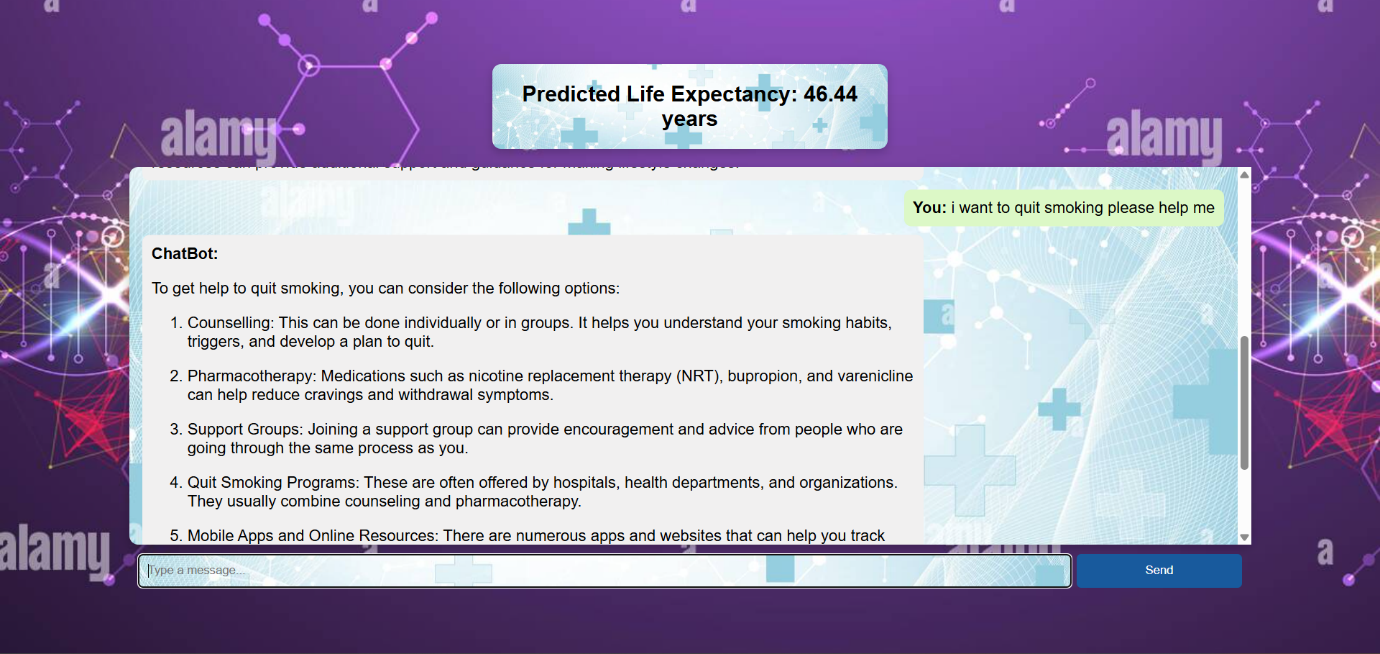
****

Figure 4.7: Ongoing Chat with the AI

* User follow-up: “I want to quit smoking please help me”
* Bot response: “1. Counselling: This can be done individually or in groups…”
* *Purpose:* Shows memory-aware, dynamic conversation using the LLM.

**4.3 Significance of the Results Obtained**

The results obtained across the model output, vector indexing, and live AI chatbot interactions validate the strength of the integrated system. The high prediction accuracy of the XGBoost model (R² = 0.89) confirms that lifestyle and health factors can be used to estimate life expectancy reliably. The preview of the custom dataset illustrates a well-structured foundation tailored specifically for this prediction task, something not available through standard datasets.

The visualization of the Pinecone vector index signifies a successful implementation of a scalable semantic retrieval system. The embedding-based approach enhances the contextual grounding of chatbot replies. The real-time personalized recommendation generation and memory-aware chat demonstrate that the system goes beyond static responses, offering tailored health guidance — a major leap toward making AI a digital wellness companion.

The seamless user interaction — from form submission to chatbot engagement — also reflects on the project’s usability and readiness for real-world application. Together, these elements highlight the platform’s ability to offer preventive, proactive, and personalized health insights.

**4.4 Conclusion**

The result analysis affirms that the system is accurate, responsive, and user-centric. From dataset creation to ML modeling and LLM-driven personalized advice, each module has worked cohesively to meet the project objectives. The prediction system correctly estimates life expectancy with low error margins, while the AI recommendation engine delivers insightful, evidence-based health suggestions through an intuitive chat interface.

Incorporating features like memory tracking, semantic retrieval, and input validation has enhanced the system's trustworthiness and usability. These outcomes strongly indicate that the platform is not only technically robust but also aligned with modern healthcare goals — prevention, personalization, and patient engagement.

Future enhancements such as integrating IoT-based real-time sensors and mobile app deployment will further increase its practical utility and impact in digital health.

**CHAPTER 5**

**CONCLUSION AND FUTURE SCOPE OF WORK**

**5.1 Work Conclusion**

This project aimed to build an intelligent system for assessing life expectancy and providing personalized health recommendations using AI and IoT-inspired data. The central idea was to combine structured health and lifestyle data with machine learning for prediction, and natural language models for interactive, contextual health guidance. A custom dataset was created due to the unavailability of publicly suitable data, and it was validated against known medical insights.

An XGBoost regression model was trained on this dataset to predict life expectancy with strong accuracy. In parallel, a domain-specific knowledge base was embedded and connected to a Mistral-7B large language model using Retrieval-Augmented Generation (RAG) via Pinecone. This integration allowed for dynamic, human-like health conversations. A full-stack web interface using HTML, JavaScript, and FastAPI was developed to deliver the predictions and chatbot experience in an accessible manner.

The results showed that combining predictive modeling with LLM-based advice leads to an engaging, trustworthy, and informative platform. The system supports multi-turn dialogue, remembers user history, and generates relevant suggestions based on form input and query context — effectively acting as a digital wellness assistant.

**5.2 General Conclusions**

The project successfully demonstrated how artificial intelligence — when combined with structured health data and natural language systems — can be used to build intelligent, proactive health guidance platforms. The use of XGBoost enabled accurate lifespan predictions, while the integration of LLMs allowed for personalized, context-rich health suggestions. The chatbot not only provides advice but also continues the conversation meaningfully, adapting to user behavior over time.

The custom dataset creation process ensured that relevant features such as stress, sleep, social interaction, smoking, and physical activity were incorporated — all of which significantly influence lifespan. The platform is designed for modularity and scalability, making it adaptable for future extensions such as real-time sensing and behavioral tracking.

This work not only meets its technical goals but also aligns with the global shift toward preventive and personalized healthcare. The results prove that such a system can contribute meaningfully in digital health spaces like virtual health assistants, fitness coaching, or chronic disease management platforms.

**5.3 Future Scope of Work**

*5.3.1 Real-Time IoT Data Integration***:** The next logical step is the integration of IoT-based health metrics such as heart rate, body temperature, SpO₂, and step count using wearable sensors or microcontrollers like Raspberry Pi. This would make the system more dynamic and reflective of a user’s current physical state, improving the relevance and timeliness of both prediction and advice.

*5.3.2 Mobile App Deployment and Notifications:* Building a mobile app version of the platform would greatly improve user accessibility and adoption. Features such as daily health insights, reminders, personalized goal tracking, and push notifications can be introduced to enhance long-term engagement and adherence to healthy habits.

*5.3.3 Integration with Electronic Health Records (EHRs):*In the future, the system can be enhanced by integrating with Electronic Health Records (EHRs) to access clinically verified user data such as medical history, prescriptions, past diagnoses, and lab reports. This would allow the AI to generate even more precise and medically informed recommendations. It would also reduce manual data entry, improve clinical relevance, and support the system's use in real-world healthcare environments such as hospitals, clinics, or telemedicine platforms

**REFERENCES**

[1] Worldometers, “Life Expectancy by Country and in the World (2025),” *Worldometers*. [Online]. Available: <https://www.worldometers.info/demographics/life-expectancy/>. [Accessed: Apr. 22, 2025].

[2] X. W. Wen, C. Y. Tsai, and J. S. Lee, “Minimum amount of physical activity for reduced mortality and extended life expectancy: a prospective cohort study,” *The Lancet*, vol. 378, no. 9798, pp. 1244–1253, Oct. 2011. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(11)60749-6/fulltext.

[3] R. L. Reimers, “Does Physical Activity Increase Life Expectancy? A Review of the Literature,” *Journal of Aging Research*, vol. 2012, Article ID 243958, 2012. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3395188/>.

[4] World Health Organization, “Physical activity,” *WHO Fact Sheets*. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/physical-activity>. [Accessed: Apr. 22, 2025].

[5] Wikipedia contributors, “World population,” *Wikipedia, The Free Encyclopedia*. [Online]. Available: <https://en.wikipedia.org/wiki/World_population>. [Accessed: Apr. 22, 2025].

[6] M. F. Springmann et al., “Life expectancy can increase by up to 10 years following sustained dietary improvements,” *Nature Food*, vol. 4, pp. 1–10, 2023. [Online]. Available: <https://www.nature.com/articles/s43016-023-00868-w>.

[7] Statistics Netherlands (CBS), “Heavy smokers cut their lifespan by 13 years on average,” *CBS News*, Sep. 15, 2017. [Online]. Available: <https://www.cbs.nl/en-gb/news/2017/37/heavy-smokers-cut-their-lifespan-by-13-years-on-average>.

[8] Harvard Health Publishing, “Sorting out the health effects of alcohol,” *Harvard Health Blog*, Aug. 6, 2018. [Online]. Available: <https://www.health.harvard.edu/blog/sorting-out-the-health-effects-of-alcohol-201808061667>.

[9] Mayo Clinic Staff, “How many hours of sleep are enough for good health?,” *Mayo Clinic*. [Online]. Available: <https://www.mayoclinic.org/healthy-lifestyle/adult-health/expert-answers/how-many-hours-of-sleep-are-enough/faq-20057898>. [Accessed: Apr. 22, 2025].

[10] M. R. Frone, “The effect of job stress on smoking and alcohol consumption,” *Journal of Occupational Health Psychology*, vol. 6, no. 4, pp. 267–277, 2001. [Online]. Available: <https://www.researchgate.net/publication/230564274_The_effect_of_job_stress_on_smoking_and_alcohol_consumption>.

[11] Harvard T.H. Chan School of Public Health, “The importance of connections: Ways to live a longer, healthier life,” *Harvard T.H. Chan School of Public Health News*, Jan. 24, 2024. [Online]. Available: <https://www.hsph.harvard.edu/news/the-importance-of-connections-ways-to-live-a-longer-healthier-life/>.

[12] E. Kavlakoglu and E. Russi, “What is XGBoost?,” *IBM Think*, IBM. (Web reference)

[13] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, USA, Aug. 2016, pp. 785–794.

[14] D. L. Longo, A. S. Fauci, D. L. Kasper, S. L. Hauser, J. L. Jameson, and J. Loscalzo, *Harrison's Principles of Internal Medicine*, McGraw-Hill Education, 18th ed., ISBN: 9780071748896.

[15] I. Wilkinson, T. Raine, K. Wiles, and G. Goodhart, *Oxford Handbook of Clinical Medicine*, Oxford University Press, 10th ed., ISBN: 9780199689903.

[16] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” *arXiv preprint*, arXiv:1908.10084, 2019. [Online]. Available: <https://arxiv.org/abs/1908.10084>.

[17] Pinecone Systems Inc., “Pinecone Documentation: Vector Database for Semantic Search and RAG,” *Pinecone*. (Web reference)

[18] National Institutes of Health (NIH), “PubMed Central – Biomedical and Life Sciences Journal Literature,” *PubMed Central*. (Web reference)

[19] M. Kumar et al., “Healthcare Internet of Things (H-IoT): Current Trends, Future Prospects, Applications, Challenges, and Security Issues,” *Electronics*, vol. 12, no. 9, p. 2050, 2023. [Online]. Available: <https://doi.org/10.3390/electronics12092050>.

[20] A. Kumar, A. De, and R. Gill, “Internet of things in healthcare: Technologies, applications, opportunities and challenges,” *AIP Conference Proceedings*, vol. 2495, no. 1, p. 020043, 2023. [Online]. Available: <https://doi.org/10.1063/5.0123090>.

[21] S. M. Yassin et al., “A Survey on IoT Applications in Health Care and Challenges,” *Proceedings of International Conference on Communication and Computational Technologies*, Springer, Singapore, 2021, pp. 343–351. [Online]. Available: <https://doi.org/10.1007/978-981-16-3246-4_31>.

[22] L. Gaur, “Internet of Things in Healthcare,” in *Geospatial Data Science in Healthcare for Society 5.0*, Springer, Singapore, 2022, pp. 97–108. [Online]. Available: <https://doi.org/10.1007/978-981-16-9476-9_6>.

[23] A. Kislay et al., “A Review on Internet of Things in Healthcare Applications,” in *Cognitive Informatics and Soft Computing*, Springer, Singapore, 2022, vol. 375, pp. 379–388. [Online]. Available: <https://doi.org/10.1007/978-981-16-8763-1_31>.

[24] S. Ghanavati et al., “Cloud-assisted IoT-based health status monitoring framework,” *Cluster Computing*, vol. 20, pp. 1843–1853, 2017. [Online]. Available: <https://doi.org/10.1007/s10586-017-0847-y>.

[25] F. Khan, “Emerging Trends and Challenges of IoT in Smart Healthcare Systems, Smart Cities and Education,” *Sensors*, vol. 24, no. 17, p. 5735, 2024. [Online]. Available: <https://doi.org/10.3390/s24175735>.

[26] P. P. Singh, “IoT in Healthcare: Opportunities and Challenges,” *International Transactions in Artificial Intelligence*, vol. 8, no. 8, 2024. [Online]. Available: <https://isjr.co.in/index.php/ITAI/article/view/228>.

[27] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proceedings of the 22nd ACM SIGKDD*, 2016, pp. 785–794.

[28] Mistral AI, “Open LLM Leaderboard,” *HuggingFace*. (Web reference)

[29] Pinecone Blog, “Why Use a Vector Database?” *Pinecone Docs*. (Web reference)

[30] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,” *arXiv preprint*, arXiv:1908.10084. [Online]. Available: <https://arxiv.org/abs/1908.10084>.

[31] LangChain, “LangChain Documentation and Use Cases,” *LangChain*. (Web reference)

[32] FastAPI Docs, “FastAPI Framework,” by Sebastián Ramírez. (Web referencePROJECT DETAILS

|  |  |  |  |
| --- | --- | --- | --- |
| *Student Details* | | | |
| **Student Name** | **Hardik Sen** | | |
| Register Number | 210907288 | Section / Roll No | A |
| Email Address | hardik.sen@learner.manipal.edu | Phone No (M) | 8114494109 |
|  | | | |
| *Project Details* | | | |
| **Project Title** | **Healthwise Lifespan Assessment Using IoT and AI** | | |
| Project Duration | 6 Months | Date of reporting | 3rd Jan 2025 |
| Expected date of completion of project | 30th June 2025 |  |  |
|  |  | | |
| *Organization Details* | | | |
| **Organization Name** | **Manipal Institute of Technology, Manipal** | | |
| Full postal address with pin code | Manipal Institute of Technology, Eshwar Nagar, Udupi - Karkala Rd, Manipal, Karnataka 576104 | | |
| Website address | www.manipal.edu/mit.html | | |
|  |  | | |
|  |  | | |
| *Internal Guide Details* | | | |
| **Faculty Name** | **Prof. Prashant M Prabhu** | | |
| Full contact address with pin code | Dept. of E&C Engg., Manipal Institute of Technology, Manipal – 576 104 (Karnataka State), INDIA | | |
| Email address | prashant.prabhu@manipal.edu | | |
| *Co- Guide Details(if any)* | | | |
| **Faculty Name** | **Dr. Spoorthi Singh** | | |
| Full contact address with pin code | Dept. of Mechatronics Engg., Manipal Institute of Technology, Manipal – 576 104 (Karnataka State), INDIA | | |
| Email address | spoorthi.shekar@manipal.edu | | |