GPT-Based AI for Disease Prediction and Healthcare Advice [CURA GPT]

GROUP 4
ALAN ACHANKUNJU (CEC21CS016)
CHRIS GEORGE (CEC21CS038)
CHRISTIN JOSEPH (CEC21CS039)
DANIEL MATHEW JOSEPH (CEC21CS041)

Guided By : ANITHA M.A
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
COLLEGE OF ENGINEERING CHERTHALA

January 6, 2025

Overview

INTRODUCTION

- Artificial Intelligence (AI) is revolutionizing healthcare by enabling accessible, personalized health assessments. A crucial area within AI health applications is providing timely and accurate health assessments through conversational interfaces.
- Using a fine-tuned GPT-2 model for health advice and LSTM/RNN for the accurate disease prediction.
- Additionally, an integrated appointment scheduler allows users to seamlessly transition from AI guidance to in-person medical care .

PROBLEM STATEMENT

- The challenge is to develop a reliable, chat-based AI system that provides accurate disease predictions and personalized health advice based on user-inputted symptoms.
- Conversational platform where users can receive health insights and schedule appointments with healthcare providers, bridging the gap between AI guidance and professional medical consultation.

OBJECTIVES

- To develop a chat-based AI capable of interpreting user symptoms through natural language input, Using fine-tuned GPT-2 and LSTM/RNN for disease prediction and health advices.
- Enable automated appointment scheduling within the platform, allowing users to select doctors, view available time slots, and secure appointments for follow-up consultations.

1.A Symptom Evaluation System on Medical Diagnosis[1]

- **Authors**: Berkay Murat, Arda Ogulcan Uzer, Sarp Ketenci, Sinan Yas, bek and Ilker Korkmaz
- **Year of Publication**: 2021
- **Key Points:**
  - The paper primarily discusses the design and implementation of the Symptom Evaluation System (SES), a software application aimed at improving the efficiency of medical consultations
  - SES allows patients to input their symptoms and relevant information before their appointments, enabling doctors to review this data in advance.

Advantages:

- Allows precise symptom input with 3D body model visualization, helping doctors understand patient issues more clearly.
- Cross-platform access via Flutter and Go ensures a responsive experience on both web and mobile.

Disadvantages:

- Relies on patient accuracy for symptom reporting, which may lead to misunderstandings in diagnosis.
- Lacks automated diagnostic support for patients, offering no preliminary health insights before seeing a doctor.

Future Scope:

- Integration of AI for Big Data:
- Expansion of 3D Modeling Features

2. Demonstration and Validation of an Advanced Symptom Checker[2]

- **Author**:Ananda Perera
- **Year of Publication**: 2021
- **Key Points**:
  - The paper discusses the development and validation of an advanced symptom checker named CAMEOS-CHECKER.
  - The system was validated against the ISABEL symptom checker, a well-known tool, to ensure accuracy and reliability.
  - The results show that CAMEOS-CHECKER performs comparably to ISABEL, particularly in sensitivity, positive predictive value, and overall accuracy

LITERTAURE SURVEY

- **Advantages:**
  - Streamlined symptom selection using a pre-set options box enhances ease of use and reduces data entry errors
  - Modular design with separate input, inference, and output components allows for easy updates and optimized performance
- **Disadvantages:**
  - Diagnostic capabilities are limited by the knowledge base, potentially missing rare or specialty-specific conditions
  - Low-probability diagnoses may result in no output or unreliable suggestions, impacting user confidence
- **Future Scope:**
  - Integration with Healthcare Systems
  - Global Deployment

LITERATURE SURVEY

3.Chatbot-based Disease Prediction and Treatment Recommendation using AI[3]

- **Author:** Chandramaprasad Pathak, Namrata Ansari
- **Year of Publication:** 2021
- **Key Points:**
    - The paper proposes a chatbot-based system for disease prediction and treatment recommendation using AI. The method involves: NLP, KNN, Dialogflow API, Training Models
    - The system uses KNN for disease prediction, which is effective for pattern recognition and offers high accuracy in medical diagnosis.
    - Dialogflow is employed to enhance the chatbot's conversational abilities, making the interaction more natural and efficient.

Literature Survey

Advantages:

- Provides easy access to healthcare information by allowing users to receive disease predictions and treatment recommendations remotely.
- Integrates K-Nearest Neighbor (KNN) for symptom analysis, improving prediction accuracy through a widely used and reliable algorithm.

Disadvantages:

- Relies on accurate symptom input by the user, which may lead to incorrect predictions if symptoms are poorly described.
- Limited by the pre-defined symptom data in its training set, potentially missing rare or less common diseases.

Future Scope

- Expansion of Disease Coverage
- Integration with Wearable Devices

4.AI-Based Medical Chatbot for Disease Prediction[4]

- **Author:** Ashish Zagade, Vedant Killedar, Onkar ManeGanesh, Nitalikar, Smita Bhosale
- **Year of Publication:** 2022
- **Key Points:**
  - The research focuses on developing an AI-based medical chatbot that leverages machine learning and NLP to predict diseases and provide healthcare support.
  - The chatbot can handle a large volume of interactions, making it a scalable solution for addressing healthcare accessibility issues.
  - The chatbot utilizes advanced technologies including AI, machine learning, and NLP to understand user queries and offer accurate medical advice.

LITERATURE SURVEY

- **Advantages:**
  - Utilizes NLP to interpret user queries accurately, enhancing user experience and comprehension.
  - Provides information on disease prevention and hospital resources, which promotes public health awareness.
- **Disadvantages:**
  - Focuses mainly on infectious diseases, which limits its application for broader health conditions.
  - Relies heavily on user-input data, which could lead to inaccuracies if input is unclear or incorrect.
- **Future Scope:**
  - Enhanced Disease Prediction Models
  - Integration of Real-time Data

LITERATURE SURVEY

5.An AI-Based Medical Chatbot Model for Infectious Disease Prediction[5]

- **Author:** SANJAY CHAKRABORTY, , HRITHIK PAUL , SAYANI GHATAK , SAROJ KUMAR PANDEY, ANKIT KUMAR, KAMRED UDHAM SINGH, MOHD ASIF SHAH
- **Year of Publication:** 2023
- **Key Points:**
  - The chatbot uses LSTM and RNN models for processing sequential data and making accurate predictions about infectious diseases.
  - Decision Tree models are used to enhance the decision-making process of the chatbot.
  - The chatbot is not only a medical tool but also a resource-saving measure for healthcare systems by automating routine tasks.

LITERATURE SURVEY

- **Advantages:**
  - Achieves high prediction accuracy of 94.32 % by using LSTM and RNN, making it reliable for infectious disease detection
  - Provides multi-modal interaction (text and voice) for improved accessibility and user experience.
- **Disadvantages:**
  - Focuses only on COVID-19 and a few diseases, which restricts its utility for broader healthcare needs.
  - Requires constant updates to maintain relevance, which can be challenging and resource-intensive.
- **Future Scope:**
  - Cross-Platform Accessibility
  - Real-time Data Integration

6.Chatbot For Disease Prediction And Treatment Recommendation[6]

- **Author:** Mrs. Swathi Sa , Sushruth Sb , Rajamani Rc , Manjunath S d , Ullas Re
- **Year of Publication:** 2023
- **Key Points:**
  - To develop a medical chatbot using machine learning and NLP that can predict diseases based on symptoms and recommend treatments.
  - The system uses technologies like Jupyter, Python, and PyCharm for development, and algorithms such as Support Vector Machine (SVM) and Naive Bayes for disease prediction.
  - The chatbot interacts with users through a user-friendly interface, processes their input using machine learning models, and retrieves relevant medical information from a pre-trained database to provide responses.

LITERATURE SURVEY

- **Advantages:**
  - Provides easy access to medical consultation through a chatbot, reducing the need for hospital visits.
  - Uses natural language processing (NLP) for user-friendly interaction, allowing users to communicate symptoms conveniently.
- **Disadvantages:**
  - Limited to identifying diseases based on predefined data, which may limit accuracy for uncommon symptoms.
  - Reliant on user-provided information, which can affect prediction accuracy if symptoms are not described well.
- **Future Scope:**
  - EIntegration with Wearable Devices
  - Enhanced AI Capabilities

LITERATURE SURVEY

7.Chatbot based Disease Prediction Using machine learning[7]

- **Author:** V. Sai Susmita, N.N.G. Tarun, D. Adhitya Kalyan, D. Dharani, V.S.V.S. Murthy
- **Year of Publication:** 2023
- **Key Points:**
  - The system uses KNN for disease prediction, which is effective for pattern recognition and offers high accuracy in medical diagnosis.
  - The approach emphasizes accessibility and convenience, aiming to provide a portable and easy-to-use solution for healthcare.

LITERATURE SURVEY

- **Advantages:**
  - Uses NLP to interpret symptoms and provide relevant health responses, enhancing user engagement.
  - Offers food and physical activity suggestions alongside disease prediction, promoting holistic health management.
- **Disadvantages:**
  - Limited to the accuracy of the Stochastic Gradient Descent algorithm, which may not perform well on diverse symptoms.
  - Relies on user-provided symptom descriptions, which can reduce prediction accuracy if inputs are unclear.
- **Future Scope:**
  - Expansion of Disease Coverage
  - Integration with Wearable Devices

Literature Survey

Name	Author(s)	Year	Advantages	Disadvantages	Method
A Symptom Evaluation System on Medical Diagnosis	Berkay Murat, Arda Ogulcan Uzer, Sarp Ketenci, Sinan Yas, bek and Ilker Korkmaz	2021	Allows precise symptom input with 3D body model visualization	Relies on patient accuracy for symptom reporting, which may lead to misunderstandings in diagnosis	Evaluation System (SES)
Demonstration and Validation of an Advanced Symptom Checker	Ananda Perera	2021	Streamlined symptom selection using a pre-set options box	Diagnostic capabilities are limited by the knowledge base	Source Separation and Voice Conversion

Literature Survey

Name	Author(s)	Year	Advantages	Disadvantages	Method
Chatbot-based Disease Prediction and Treatment Recommendation using AI	Namrata Ansari	2021	Integrates K-Nearest Neighbor (KNN) for symptom analysis	Limited by the pre-defined symptom data in its training set	NLP, KNN, Dialogflow API
AI-Based Medical Chatbot for Disease Prediction	Ashish Zagade	2022	Utilizes NLP to interpret user queries accurately,	Focuses mainly on infectious disease	

Literature Survey

Name	Author(s)	Year	Advantages	Disadvantages	Method
An AI-Based Medical Chatbot Model for Infectious Disease Prediction	SANJAY VARMA	2023	Achieves high prediction accuracy of 94.32	Requires constant updates to maintain relevance	LSTM AND RNN

Conclusion for Literature Survey

- These studies collectively demonstrate that AI in healthcare, especially chatbots and symptom checkers, can significantly aid early disease detection and healthcare access.
- Most systems combine machine learning (e.g., KNN, SVM, LSTM) and NLP to enhance user interaction and diagnosis.
- From the survey we understood that LSTM/RNN have better accuracy than other.
- The use of LSTM and RNN models demonstrates strong performance in processing sequential medical data, achieving high accuracy in disease prediction by effectively capturing temporal patterns in symptoms.

PRODUCT FUNCTIONS

- Provides general healthcare advice through the chat, which could include suggestions for symptom management, lifestyle changes, or dietary recommendations.
- Users can request specific health-related tips, such as managing stress, reducing smoking, or handling chronic conditions.
- A dedicated section for booking appointments with healthcare providers. Users can view available time slots, select doctors, and schedule appointments, facilitating a smooth transition from digital advice to in-person consultation.

PROPOSED SYSTEM

- To develop a chat-based AI capable of interpreting user symptoms through natural language input, Using fine-tuned GPT-2 and LSTM/RNN for disease prediction and health advices.
- Healthcare Advice: Offer tailored advice, including lifestyle changes, preventive measures, and when to seek professional help.
- User-Friendly Interface: Develop an interface for easy user interaction, available via web.

Requirements

Software Requirements:

- Operating System:Linux (Ubuntu, CentOS) or Windows Server for a production environment MacOS or Windows for development environments
- Programming Languages:Python,html,css,javascript

Hardware Requirements:

- Processor: Intel Core i5 or equivalent (minimum); i7 or higher recommended
- RAM: At least 16 GB (32 GB or higher recommended for faster model training and testing)
- Storage: SSD with at least 500 GB of free space
- GPU: NVIDIA GPU (such as GTX 1080 or higher) for local training and fine-tuning

Tools

The following tools are used:

- **Development Environment:**
  - Jupyter Notebook, PyCharm for coding and experimentation.
- **Data Handling:**
  - Pandas, NumPy for data manipulation and analysis.
- **Natural Language Processing:**
  - NLTK, spaCy for text processing and feature extraction.

Languages

- **Python:**
  - For implementing natural language processing.
  - Extensive libraries and community support .
- **JavaScript/HTML/CSS:**
  - For developing the user interface .
  - Ensures a responsive and interactive user experience.

Architecture Diagram

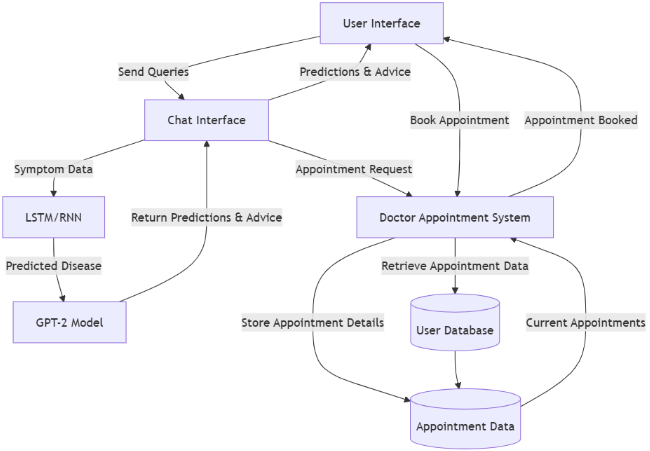

Fig : Architecture Diagram

Usecase Diagram

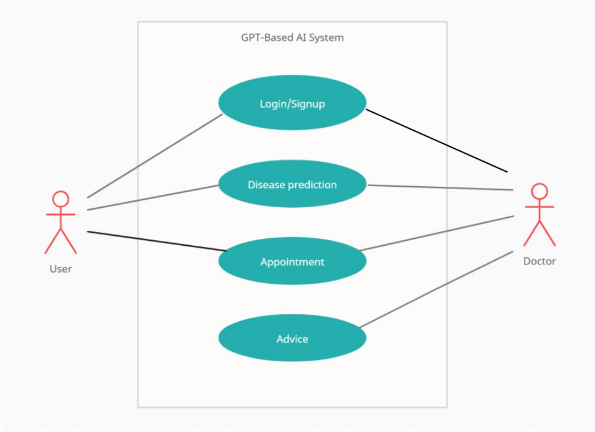

Fig : Usecase Diagram

DFD Level 0 Diagram

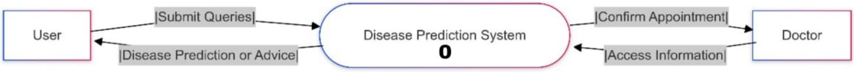

Fig : DFD Level 0 Diagram

DFD Level 1 Diagram

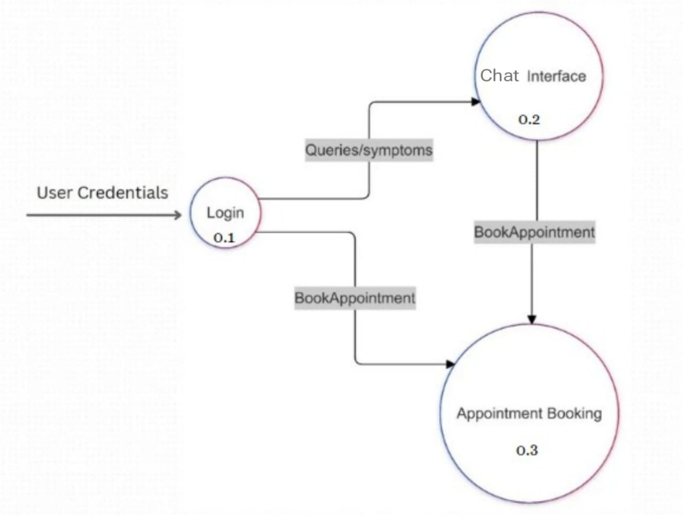

Fig : DFD Level 1 Diagram

DFD Level 2 Diagram

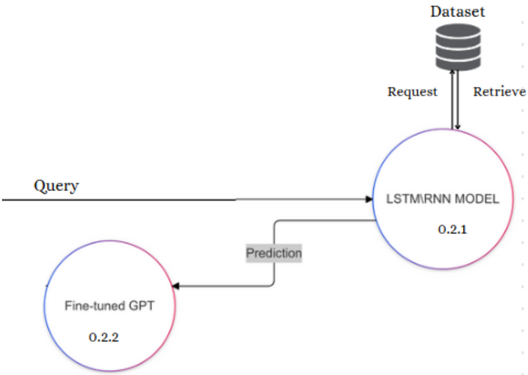

Fig : DFD Level 2 Diagram

Sequence Diagram

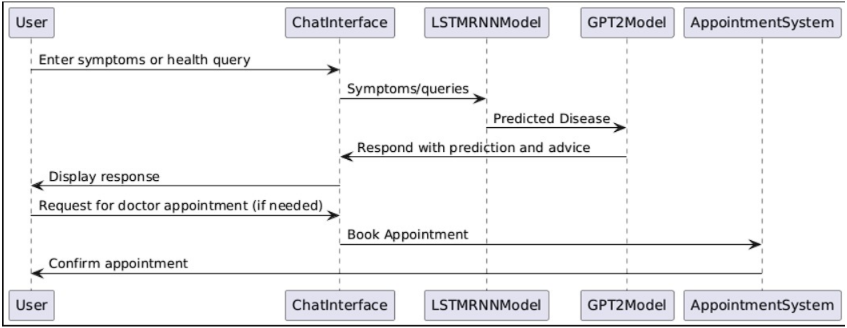

Fig : Sequence Diagram

Activity Diagram

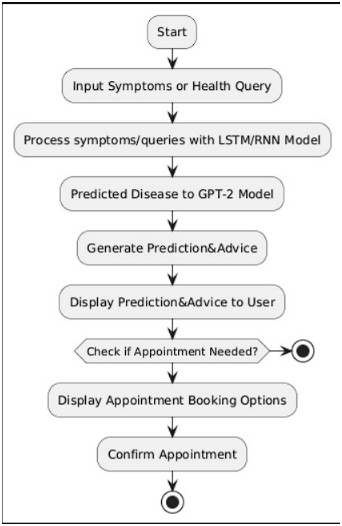

Fig : Activity Diagram

Gantt chart

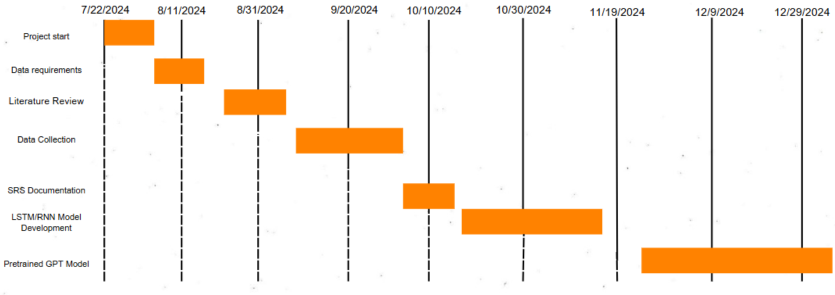

Fig : Gantt chart

COST ESTIMATION

Overview

- Cost Estimation is performed using the COCOMO Model, which helps predict:
  - **Effort** required to complete the project.
  - **Time (Months)** needed for development.
  - **Cost** based on estimated effort and developer salary.
- The project is classified as **Organic** due to its relatively simple structure and straightforward requirements.

COST ESTIMATION

Input Parameters

- Lines Of Code (LOC) = 200
- Project Type = Organic
- Salary per month = 50,000

COCOMO Model Formulas

$$Effort = a \cdot (\text{KLOC})^b \quad \text{PM}$$

$$Time = c \cdot (E)^d$$

$$\text{Cost} = \text{Effort} \cdot \text{Salary}$$

where a = 2.4, b = 1.05, c = 2.5, d = 0.38 for Organic.

COST ESTIMATION

KLOC (Kilo Lines of Code)

$$KLOC = \frac{200}{1000} = 0.2$$

Effort (E) in Person-Months (PM)

$$E = 2.4 \times (0.2)^{1.05} = 2.4 \times 0.203 = 0.49\, PM$$

Time (T) in Months

$$T = 2.5 \times (0.49)^{0.38} = 2.5 \times 0.76 = 1.9\,\text{Months}$$

Cost (C)

$$C = 0.49 \times 50,000 = 24,500$$

Final Estimates: Effort: 0.49 Person-Months

Time: 2 Months
Cost: 24,500

Implementation

```python
import pandas as pd
import re
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Embedding, Dropout
from tensorflow.keras.preprocessing.text import Tokenizer
from tensorflow.keras.preprocessing.sequence import pad_sequences
from sklearn.model_selection import train_test_split
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
import nltk

# Download nltk data files (run this once)
nltk.download('stopwords')
nltk.download('wordnet')

# Initialize lemmatizer and stopwords
lemmatizer = WordNetLemmatizer()
stop_words = set(stopwords.words('english'))

# Text preprocessing function
def preprocess_text(text):
    # Convert to lowercase
    text = text.lower()
    # Remove special characters, numbers, and punctuation
    text = re.sub(r'[^a-zA-Z\s]', '', text)
    # Tokenize and remove stopwords, apply lemmatization
    tokens = text.split()
    tokens = [lemmatizer.lemmatize(word) for word in tokens if word not in stop_words]
    return ' '.join(tokens)
```

Fig : Implementation

Implementation

```
tokenizer = Tokenizer(num_words=5000, oov_token="<OOV>")
tokenizer.fit_on_texts(data['symptoms'])
sequences = tokenizer.texts_to_sequences(data['symptoms'])
max_len = max([len(x) for x in sequences])
padded_sequences = pad_sequences(sequences, maxlen=max_len, padding='post')

# One-hot encode the target labels
disease_labels = pd.get_dummies(data['disease']).values
```

```
# Train-test split
X_train, X_test, y_train, y_test = train_test_split(padded_sequences, disease_labels, test_size=0.2, random_state=42)
```

```
# Build the LSTM model
model = Sequential()
model.add(Embedding(input_dim=5000, output_dim=64, input_length=max_len))
model.add(LSTM(128, return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(64))
model.add(Dense(32, activation='relu'))
model.add(Dense(disease_labels.shape[1], activation='softmax'))

# Compile the model
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])

# Train the model
model.fit(X_train, y_train, epochs=10, batch_size=32, validation_data=(X_test, y_test))
```

Conclusion

- The proposed chat-based AI system provides accurate disease predictions and personalized health advice based on user-inputted symptoms.
- By leveraging natural language processing capabilities, the system allows users to interact via a conversational chat interface, receive predictions on possible health conditions based on their symptoms, and gain general wellness advice

REFERENCES

1 B. Murat, A. O. Uzer, S. Ketenci, S. Yas¸bek, and I. Korkmaz, "A Symptom Evaluation System on Medical Diagnosis," 2021.

2 A. Perera, "Demonstration and Validation of an Advanced Symptom Checker," 2021.

3 C. Pathak and N. Ansari, "Chatbot-based Disease Prediction and Treatment Recommendation using AI," 2020.

4 A. Zagade, V. Killedar, O. ManeGanesh, N. Nitalikar, and S. Bhosale, "AI-Based Medical Chatbot for Disease Prediction," 2021.

5 S. Chakraborty, H. Paul, S. Ghatak, S. K. Pandey, A. Kumar, K. U. Singh, and M. A. Shah, "An AI-Based Medical Chatbot Model for Infectious Disease Prediction," 2022.

6 Chatbot For Disease Prediction And Treatment Recommendation , Mrs. Swathi Sa , Sushruth Sb , Rajamani Rc , Manjunath S d , Ullas Re , 2021

7 V. S. Susmita, N. N. G. Tarun, D. A. Kalyan, D. Dharani, and V. S. V. S. Murthy, "Chatbot based Disease Prediction Using Machine Learning," 2022.