SAT 5144 Artificial Intelligence in Healthcare

Progress Report 1

Name: Yaswanth Ganapathi, Ganesh Vannam

Group Number: 27

Title: AI-Driven Personalized Health Assistant: Adaptive Recommendations Based on Individual Health Profiles.

Introduction:

The goal of this project is to develop an AI-driven medical decision support system that assists healthcare professionals in diagnosing medical conditions and suggesting treatments. The system will analyze patient symptoms, medical histories, and other relevant factors to predict diseases and recommend suitable treatment plans.

We aim to utilize advanced AI techniques such as machine learning and natural language processing (NLP) to process and analyze medical data, ultimately improving healthcare decision-making and enhancing patient outcomes.

Literature Review:

Artificial intelligence (AI) is transforming the healthcare landscape by enabling more efficient diagnosis and treatment planning. Machine learning algorithms, particularly decision trees, random forests, and neural networks, have proven effective in predicting diseases and recommending treatments based on clinical data. These techniques help identify patterns in patient data and offer personalized recommendations, reducing human error and improving patient outcomes.

Natural Language Processing (NLP) plays a crucial role in healthcare AI by processing unstructured data from sources like medical records, clinical notes, and research articles. By leveraging NLP, AI systems can extract valuable insights, enabling more accurate disease prediction. Rajkomar et al. (2019) demonstrated the power of machine learning models trained on electronic health record (EHR) data to predict patient outcomes and assist in decision-making [1]. Integrating NLP with machine learning enhances the ability to process textual data, increasing the predictive power of diagnostic systems.

Zhou et al. (2021) illustrated the benefits of combining ML and NLP to improve diagnostic accuracy. Their study emphasized how AI-based systems could assist healthcare providers by providing accurate and timely diagnoses based on a combination of symptoms, medical history, and other clinical data. This integration of AI technologies has

significant potential to streamline healthcare processes and reduce the burden on medical professionals [2].

In a similar vein, Zhai et al. (2021) explored the integration of AI into healthcare decision-making. Their research suggests that AI systems, particularly those that combine structured and unstructured data, can provide superior diagnostic performance compared to traditional methods. This highlights the growing potential of AI in shaping the future of healthcare [3].

Dataset Description:

For this project, the **Medical Recommendation Dataset** (medical_data) from Kaggle has been chosen a for the proof of concept to ensure the feature is working, which includes data on patient symptoms, diagnosed diseases, medical histories, and prescribed treatments. Later we will use the Patient health record data for the feature.

- **Features**:
  - **Name**: Patient's name
  - **DateOfBirth**: Date of birth
  - **Gender**: Gender of the patient
  - **Symptoms**: Reported symptoms
  - **Causes**: Underlying causes of symptoms
  - **Disease**: Diagnosed disease
  - **Medicine**: Prescribed treatment

Dataset Size: Approximately 500 records, with missing values handled appropriately.
Dataset Link: https://www.kaggle.com/datasets/joymarhew/medical-reccomadation-dataset

Methodology :

The project aims to build a medical decision support system capable of predicting diseases and recommending treatments. The methodology consists of several stages, including data preprocessing, model development, and evaluation.

Steps Taken:

1. **Data Preprocessing**:
   a. **Symptom Tokenization**: Convert the symptoms field into a format suitable for NLP models by tokenizing the text and removing stopwords.
   b. **Feature Engineering**: Additional features like **Age** (derived from **DateOfBirth**) and **Gender** will be included to enrich the model.

c. **Categorical Encoding**: Convert categorical features such as **Disease** and **Gender** into numerical values using techniques like LabelEncoding or One-Hot Encoding.

2. **Model Development**:
    a. The primary machine learning models to be developed include **Random Forests** and **Decision Trees**, which are suitable for classification tasks such as disease prediction.
    b. **Natural Language Processing (NLP)** will be integrated to analyze the **Symptoms** text data and improve the model's ability to handle unstructured data.
    c. **Speech Recognition** and **Text-to-Speech (TTS)** will be incorporated to allow healthcare professionals to interact with the system using voice commands.
3. **Model Evaluation**:
    a. The models will be evaluated using performance metrics such as **Accuracy**, **Precision**, **Recall**, and **F1-Score**. Cross-validation will be applied to assess the model's generalizability.
4. **Speech Interface**:
    Implement a **speech recognition** and **text-to-speech** system using the speech_recognition and pyttsx3 libraries to allow healthcare professionals to interact with the system through voice commands. The system will process spoken symptoms and generate treatment recommendations based on the data.

Future Work:

So far, our work has been focused on the development of the code behind the clinical assistant feature, ensuring its functionality. To validate our progress, we have used the **medical_data** dataset and **OpenAI** models to demonstrate the capabilities of the system, such as analyzing symptoms and suggesting treatments.

Moving forward, our goal is to integrate **patient health record data** to make the system more robust and patient-specific. This will include using patient medical histories, vital signs, and other health information to enhance the accuracy and relevance of the system's recommendations.

Additionally, we plan to incorporate the **LLaMa2** model, which will allow us to significantly improve the AI's natural language processing capabilities. This will enable the system to offer more refined and interactive conversations, transforming it into a virtual nurse capable of answering patient queries, providing insights, and assisting healthcare professionals in making better clinical decisions.

Output:

```
Listening...
You said: hi clinical assistant I'm yashwanth how are you
I'm doing well, thank you for asking! How can I assist you today?
Listening...
You said: how are my vitals today
Your vitals are: Heart Rate = 72, Sleep Hours = 6.
Listening...
You said: I'm feeling tired
I'm sorry to hear that you're feeling tired. Are you also experiencing any other symptoms?
Listening...
You said: no but I'm having a little bit of headache and fatigue
It sounds like you might be stressed. Try taking a break and drinking some water.
Listening...
You said: oh thank you exit
```

References:

1. Rajkomar, A., Oren, E., Chen, K., Dai, A. M., & Hajialiasghari, F. (2019). Scalable and accurate deep learning for electronic health records. *npj Digital Medicine*, 2(1), 1-10.
2. Zhou, X., Yang, Q., & Zhang, Y. (2021). A deep learning-based medical decision support system for disease diagnosis and treatment recommendations. *BMC Medical Informatics and Decision Making*, 21(1), 1-9.
3. Zhai, X., Li, Z., & Zhang, S. (2021). The impact of artificial intelligence on healthcare services: A systematic review. *International Journal of Environmental Research and Public Health*, 18(1), 271.

Conclusion:

This progress report highlights the development of an AI-driven clinical assistant that analyzes symptoms and recommends treatments using the medical_data dataset and OpenAI models. The system has shown promise in enhancing healthcare decision-making by providing valuable insights.

Additionally, incorporating the LLaMa2 model will further enhance the system's natural language processing capabilities, enabling it to engage in more interactive and meaningful conversations with healthcare professionals. These improvements aim to create a virtual nurse that can assist healthcare professionals in making better decisions and ultimately improve patient care through data-driven, personalized insights.