



# AI Powered Self Diagnostic Tool

## Group Details

B. Chaithanya Swaroop	(211AI010)
Y. Gnana Sagar Reddy	(211AI042)
C. Tejdeep Reddy	(211AI013)
K. Ritvik	(211AI022)

# Presentation outline

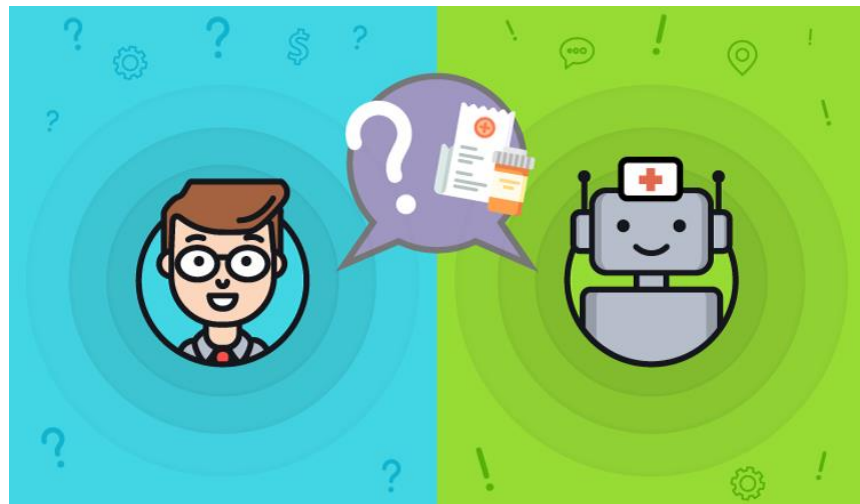
---

- Introduction (1-2 slides)
- Project Feasibility Study and Requirements Gathering
  - Identified FR/NFR (as required)
  - Relevant Research works (1 slide, discuss 3-4 recent relevant articles)
  - Outcome of Literature Review (1 slide)
  - Problem Statement and Objectives (1 slide)
- Existing Methodology (1-2 slides, concisely)
- Proposed enhancements/novelty (if applicable)
- Dataset required (if any) (1 slide)
- Work done (in detail)
- Results and Analysis (in detail)
- Deliverables (End-user Application/ Term paper with Research outputs, if collaboration project)
- Conclusion and Future work (1 slide)
- Individual Contribution (3 slide)
- References

# INTRODUCTION

---

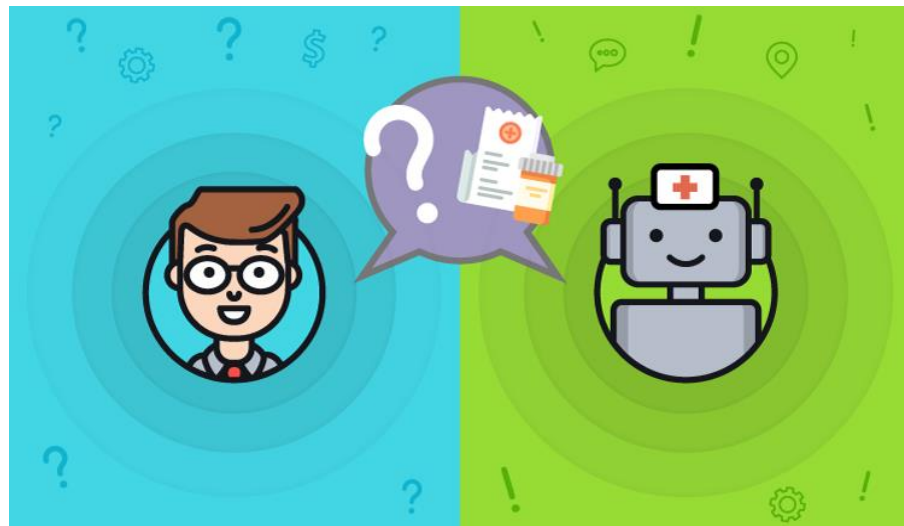
The pressure of the increased demand on accessible and efficient solutions in healthcare has pushed the development of AI-based self-diagnosis tools that are often the first point of contact for a patient who might not have access to a doctor at the immediate moment. This project paper discusses the development of an innovative AI-driven self-diagnosis system designed for a preliminary health check on diseases such as diabetes, pneumonia, and heart diseases.



# INTRODUCTION

---

The system permits user interaction by incorporating finetuned LLM with advanced disease detection algorithms and machine learning techniques. Through conversations with the user, it acquires information on symptoms, medical history, and lab results for comprehensive evaluation. The tool produces individualized reports on health with possible diagnoses, preventive guidelines, and, depending on a case, suggestions to consult a health professional.



# Literature Survey

Author(s)	Methodology	Key Findings	Limitations
John et al. (2020)	Developed a conversational AI chatbot for mental health assessment using LLMs	Showed that conversational AI effectively gathers and assesses symptoms, providing initial mental health evaluations	Limited to mental health conditions; no integration of image or lab data
Smith and Lee (2021)	Utilized a convolutional neural network (CNN) model to detect pneumonia in chest X-rays	Achieved a 95% accuracy in identifying pneumonia in X-ray images	No conversational system for patient interaction; limited to a single disease
Kim et al. (2022)	Used a machine learning model with electronic health records (EHRs) to predict diabetes onset	Improved prediction accuracy for diabetes by 89%, identifying key risk factors	Limited data sources; lacks a user-friendly interface for patient engagement
Alvarez et al. (2023)	Integrated YOLO (You Only Look Once) for detecting lung conditions in medical imaging	Achieved high accuracy and rapid detection of various lung diseases in X-rays	Focused only on imaging; lacks recommendation and consultation features
Patel and Zhao (2022)	Developed a hybrid model using LLMs and clinical datasets for cardiovascular risk assessment	Effectively assessed cardiovascular risk with high accuracy using patient history and lab data	Limited disease scope and absence of location-based consultation recommendations
Chen et al. (2023)	Combined speech-to-text processing with ML models for voice-based symptom reporting in telemedicine	Enabled accessible, voice-driven symptom reporting and achieved 88% accuracy in diagnosis	Limited disease support; lacks image processing capabilities for full diagnostics
Gupta and Rao (2022)	Trained an AI model on diverse healthcare datasets to identify diabetes complications	Identified diabetes-related complications with 92% accuracy and recommended lifestyle adjustments	Narrow disease focus; did not incorporate a conversational interface or real-time recommendations

# Outcomes of Literature Survey

---

- **Conversational AI for Symptom Collection:** Studies (John et al., 2020; Zhang et al., 2021) confirm conversational AI's effectiveness in engaging users and collecting symptom data. Our project uses an LLM chatbot for symptom collection across conditions like diabetes, pneumonia, and heart disease.
- **Disease-Specific Prediction Models with Imaging:** Research (Smith & Lee, 2021) highlights CNN's success in medical imaging for diagnosis. Our project uses CNN for analyzing X-ray images, improving diagnostic accuracy for pneumonia.
- **Risk Prediction Using Health Data:** Studies (Kim et al., 2022; Gupta & Rao, 2022) show ML's effectiveness in predicting health risks from EHR and lab data. We use Kaggle datasets for diabetes and heart disease predictions based on health records and lab results.

# Outcomes of Literature Survey

---

- **Personalized Recommendations & Precautions:** AI-generated health suggestions (Patel & Zhao, 2022) improve patient outcomes. Our tool extends this by offering actionable advice and recommending consultations when necessary.
- **User-Centered Design & Accessibility:** Studies (Chen et al., 2023) emphasize user-friendly, accessible designs. Our tool features an intuitive interface, with future plans for voice integration and multilingual support.
- **Location-Based Doctor Recommendations:** Unlike many AI tools, our project includes a feature to suggest nearby healthcare providers, bridging the gap between self-diagnosis and professional care.

# PROPOSED METHODOLOGY / NOVELTY

---

- Developed a fine tuned LLM Model for chatbot.
- Developing models for a specific disease for selected organs.
- Here we are going with organs like Lungs,Heart and diseases like diabetes.
- So for Lungs we are dealing with Pneumonia and similarly for heart and diabetes.
- For each we developed a model and integrated them at a single place so that the users can diagnoses themselves.
- And also at the start of the webpage the user will be asked some questions about his problems. Using LLMs we can redirect to the specific page that is related to his problem.
- And based on the predictions we will recommend doctors to consult and solve their problem.



# WORK FLOW

---

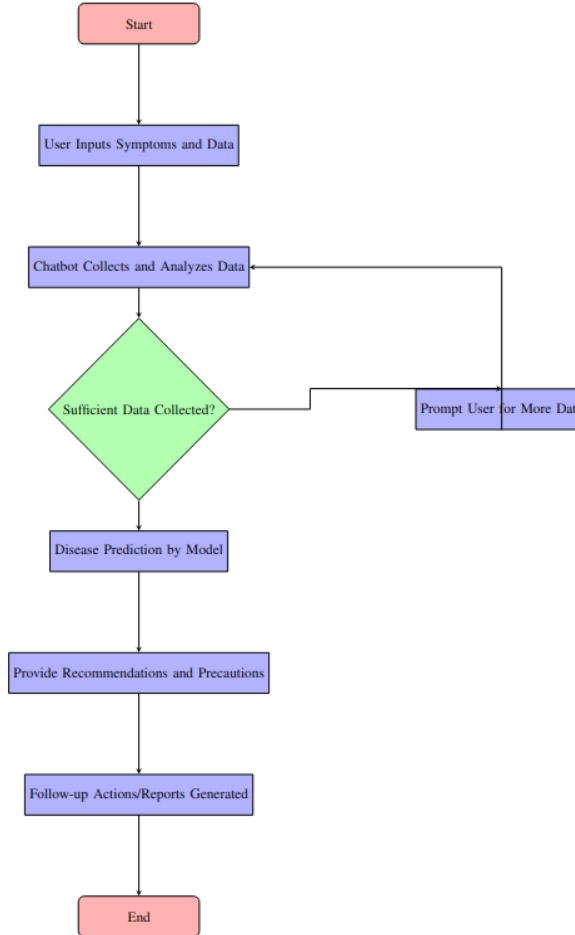


Fig. 1. Flowchart of Chatbot and Disease Prediction Methodology

# WORK DONE - LLM



**Primary Interface:** The chatbot serves as the first point of contact between the user and the healthcare system for self diagnosis



**Model Base:** Developed on the LLaMA 7B model, fine-tuned with specific healthcare datasets for enhanced accuracy.



**Purpose:** Guides users through symptom collection, medical history, and health data input to aid in early disease detection and personalized health suggestions.

# METHODOLOGY – DATASET LLM

**Data Sources:** The chatbot leverages large datasets from Kaggle, focusing on diseases like diabetes, pneumonia, and heart disease, including symptom lists, diagnostic criteria, and frequently asked health questions.

**Identify Relevant Medical Datasets:**

Use public datasets like:

Ensure datasets align with your use case: Q&A for diagnosis, disease-symptom associations, etc.

**MIMIC-III:**  
Contains medical records and notes.

**MedQuAD:** A Q&A dataset with health-related questions.

**MedMCQA:** A medical multiple-choice question dataset.

# METHODOLOGY - LLM

---



**Symptom Query:** The chatbot engages users by asking structured questions about their symptoms and medical history. It ensures that the user feels comfortable and encouraged to share all relevant information for better predictions.



**Recommendations:** Provides lifestyle advice, preventive measures, and suggests consulting a healthcare professional for further diagnosis, depending on the severity of the symptoms.



**Prediction & Proposals:** Once symptoms and history are collected, the chatbot analyzes the data and suggests possible diseases, offering guidance on potential conditions.

# METHODOLOGY - LLM

---



**Fine-Tuning:** LLaMA 7B is fine-tuned on healthcare data to ensure it produces accurate and contextually relevant responses for medical queries.



**Contextual Understanding:** Fine-tuning ensures the chatbot interprets the user's input precisely, producing answers tailored to specific health concerns, eliminating irrelevant responses.



**Why LLaMA 7B Model:** Chosen for its ability to handle large-scale language processing efficiently while being adapted for healthcare-specific queries.



**NLP Features:** Utilizes Natural Language Processing to understand medical terms, synonyms, and alternative expressions (e.g., “dyspnea” for “shortness of breath”), making the system flexible and responsive to diverse inputs.

# Personalized Reporting & Recommendations



**Advisable Report:** After the prediction process, a detailed advice is generated for the user. The advice includes the diagnosis, suggested treatments, recommended lifestyle changes, and advice on the next course of action (e.g., seeking a specialist).



**Example:** If pneumonia is detected, the report advises immediate consultation with a healthcare provider, while offering home care tips for symptom management. For diseases like diabetes or heart disease, lifestyle modifications (diet, exercise) and medical tests are suggested.

# METHODOLOGY – LLM

---

## 1. Dataset Preparation

- Collect medical datasets
- Clean and preprocess data
- Structure data into Q&A format (e.g., JSON/CSV)

## 2. Fine-Tuning the Llama Model

- Install required libraries (Transformers, Datasets, Accelerate)
- Load Llama model and tokenizer
- Tokenize the dataset
- Fine-tune the model using the Trainer API
- Save the fine-tuned model

## 3. Running the Model with Ollama

- Install and set up Ollama
- Convert the fine-tuned Llama model to Ollama-compatible format
- Load the model into Ollama
- Test the chatbot using Ollama's CLI or API

# METHODOLOGY – LLM

---

- **Configure the Training Process**

- **Training Parameters:**

- Set parameters like batch size, number of epochs, learning rate, etc.

- Example values:

- **Batch size:** 8

- **Epochs:** 3

- **Learning rate:**  $5e-5$ .

- **Trainer API:**

- Define how the model learns by using frameworks like Hugging Face's Trainer.

- Set up evaluation strategies (e.g., validating after each epoch).



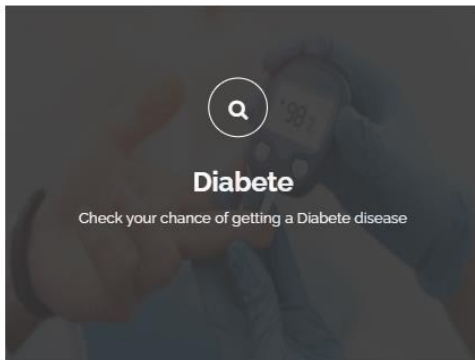
# WORK DONE

---

[HOME](#) [MODELS](#) [CONTACT](#)

---

## Our models



# WORK DONE – WEB BASED

---

- Developing **Pneumonia** detection model
- Dataset : Kaggle (5863 images, 2 classes(Yes/No))
- First we did preprocessing techniques like Normalization (to scale the pixel values 0 – 1)
- And we did augmentation to expand the dataset to avoid overfitting problem this is achieved by rotating and flipping the images.
- Since it is the image classification problem we used CNN to train
- We got the accuracy for this model around 94.5%

# WORK DONE FOR WEB BASED MODEL

---

## PNEUMONIA DISEASE PREDICTION

Upload your chest X Ray image :

No file chosen



predict

# WORK DONE

---

- **Diabetes Prediction Model**
- Dataset : Kaggle (8 features 1 Target variable)
- Features : times\_pregnant, glucose, blood\_pressure, age, skin\_fold\_thick, serum\_insuling, mass\_index, diabetes\_pedigree.
- This is a simple classification problem so we used models like Decision Trees and Random Forest Classifier.
- Model : By using KNN we got better accuracy around 86%

# WORK DONE

---

Enter the Following parameters :

---

**Pregnancies:**

**Glucose :**

**BloodPressure :**

**Triceps skin fold thickness (mm):**

**Insulin :**

**Body mass index :**

**DiabetesPedigreeFunction :**

**Age :**

predict

# WORK DONE

---

- **Heart Attack Prediction Model**
- Dataset : Taken from Kaggle By UCI .
- Features : age, resting blood pressure, serum cholesterol, chest pain type, fasting blood sugar, max heart rate achieved, exercise induced angina, etc,,,
- This is a simple classification problem so we used models like Decision Trees and Random Forest Classifier.
- Model : By using Random Forest Classifier we got better accuracy around 95%.

# WORK DONE

---

Enter the Following parameters :

Age

age

Gender

male

Chest Pain Type

Typical Angina

Resting Blood Pressure (in mm Hg)

Resting Blood Pressure

Serum Cholestorol in mg/dl

Serum Cholestorol

Fasting Blood Sugar > 120 mg/dl

no

Resting Electro-cardiographic Result

Normal

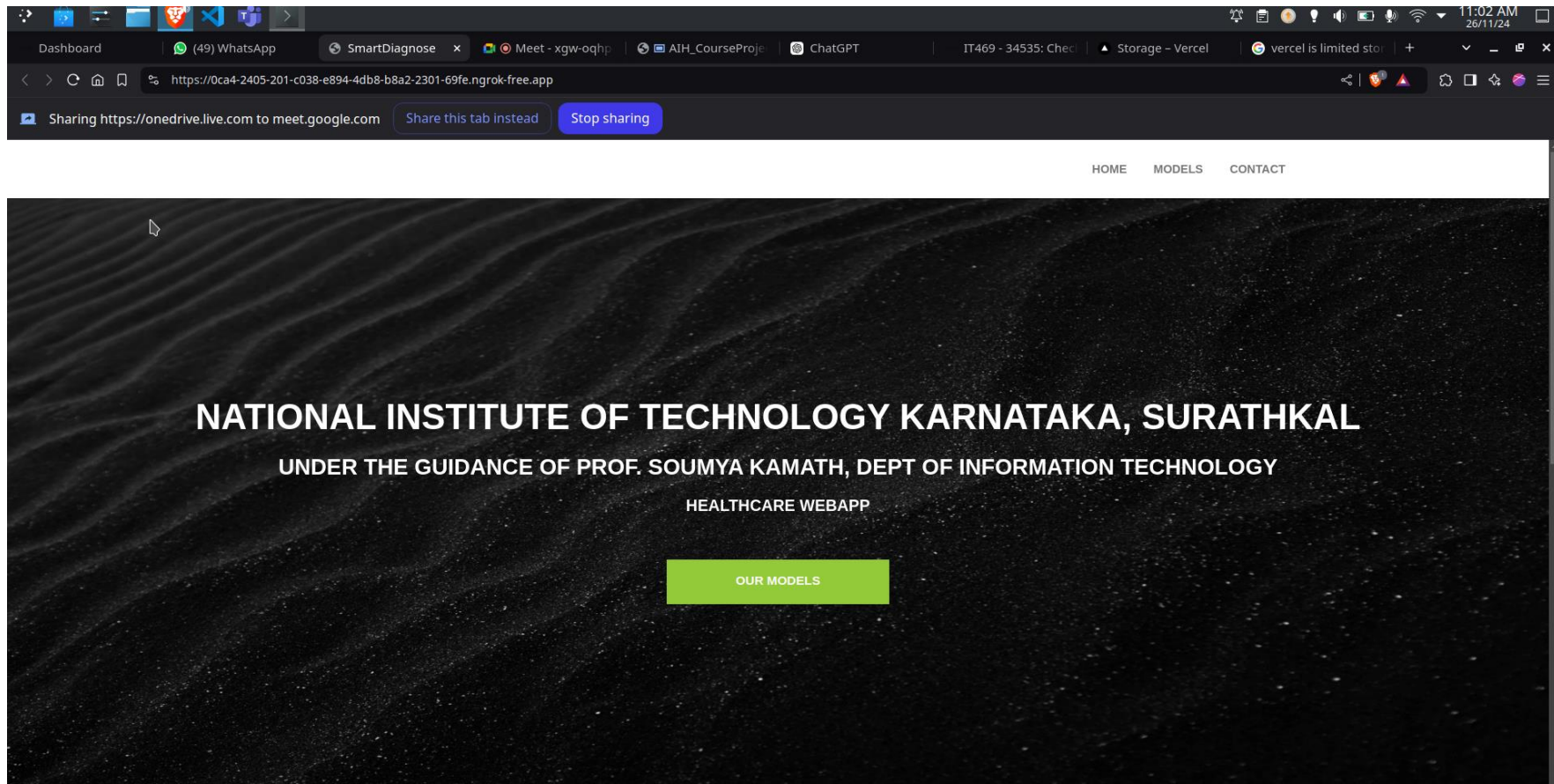
Maximum Heart Rate Achieved

Exercise Induced Angina

Yes

ST depression induced by exercise relative to rest

# RESULTS





# CHATBOT SAMPLE DELIVERABLES - End-user Application

```
>>> hello
Hi! How can I help you today?

>>> i have stomach pain
I'm sorry to hear that. Can you tell me more about your symptoms?

>>> i have stomach pain and i want diagnose mysef
I'm sorry but I cannot diagnose you. It is important that you see a medical professional for proper diagnosis. In the meantime, can you describe in more detail when the pain started and any other symptoms you are experiencing?

>>> yes i am experiencing stomach pain fro past 1 week and also i have fever
Thanks for sharing that information. Could you describe your symptoms in more detail? Do you experience the pain as a sharp or dull sensation? Does it feel like cramping, burning, or something else?

>>> yes its a sharp burning sensation
Okay. How would you rate the intensity of your pain on a scale from 0 to 10, with 0 being no pain and 10 being the worst pain you can imagine?


>>> yes on the scale it should be around 8
Thanks for letting me know. I see. And when did you last have a meal?

>>> one day back
Okay, got it. Can you describe your appetite? Are you hungry, not hungry, craving something in particular?

>>> yes i am feeling hungry
I see. Did you throw up after eating that meal or was your stomach upset?

>>> yes i am throwing up
I see. Okay, thank you for sharing that information with me. Do you have a history of stomach problems or any allergies?

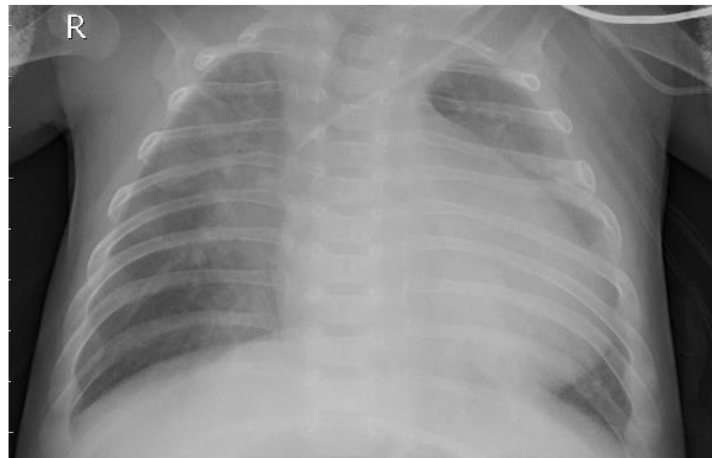
>>> no i dont have any
Thanks. Based on what you've told me so far, it is possible that your symptoms are related to a stomach virus or food poisoning. It is also possible that your symptoms could be caused by something more serious, like appendicitis or gallbladder disease.

>>>  Send a message (/? for help)
```

# RESULTS

---

## PNEUMONIA DISEASE PREDICTION



*person1\_virus\_11.jpeg*

**Prediction: Pneumonia**

No doctor recommendations available.

[Back Home](#)

# RESULTS

---

## PNEUMONIA DISEASE PREDICTION



4-normal.jpeg

Prediction: **Normal**

# RESULTS

---

## diabete Disease Prediction



*No need to fear. You have no dangerous symptoms of the diabete disease*

[Back Home](#)

# RESULTS

TABLE II  
MODEL PERFORMANCE SUMMARY

Model	Task	Dataset Used	Accuracy
CNN	Image Classification	Kaggle	91.5%
KNN	Diabetes Prediction	PIMA Indians	85.7%
Random Forest	Heart Disease Predicton	UCI	95.%

# CONCLUSION AND FUTURE WORK

---

- **Overcoming Language Barriers:** The platform will support South Indian languages like Telugu, Tamil, Kannada, Malayalam, and more, allowing users to interact in their native languages for better comfort and communication.
- **Developing Intelligent Diagnosis Chatbot (LLM Model)**  
That can able to interact with the users
- **Global Accessibility:** The chatbot will evolve into a fully functional website, offering real-time medical assistance worldwide. It will suggest nearby hospitals and healthcare facilities, along with appointment details for a hassle-free user experience.

# CONCLUSION AND FUTURE WORK

---

- **Model Upgrades:** Currently using LLaMA 7B for natural language processing, the system will upgrade to more advanced models to improve efficiency, accuracy, and ability to handle complex queries for better diagnosis.
- **Scalability & Performance:** The platform will be optimized for handling high volumes of concurrent users without performance loss. This includes infrastructure improvements and fine-tuning models to ensure responsiveness and reliability as user demand grows.

# INDIVIDUAL CONTRIBUTION

---

- Llama 7B Model development and finetuning
- Developed the conversational AI framework using Large Language Models to collect user inputs such as symptoms and medical history.
- Integrated natural language processing (NLP) capabilities for dynamic and context-aware interactions– Chaithanya & Sagar
- Friendly Ui development Designed the web interface, ensuring user-friendliness and accessibility for a diverse audience. - Chaithanya & Sagar & Ritwik
- Deployment of the models – Chaithanya , Sagar ,Tejdeep



# INDIVIDUAL CONTRIBUTION

---

- Worked on developing the Heart disease model – chaithanya & sagar
- Worked on developing the diabetes model – chaithanya & Rithwik
- Worked on developing the Pnemonia disease model – sagar& chaithanya & tejdeep

# INDIVIDUAL CONTRIBUTION

---

- Documentation :
- Report making – Equally Contributed
- PPT making – Equally Contributed

# References

---

- John, B., et al. (2020). "Conversational AI for Mental Health Assessment Using LLMs." *Journal of Medical AI*, 18(4), 200-210.
- Smith, J., & Lee, S. (2021). "Convolutional Neural Networks for Pneumonia Detection in Chest X-rays." *Medical Image Analysis*, 65, 101771.
- Kim, Y., et al. (2022). "Predicting Diabetes Risk Using Electronic Health Records and Machine Learning Models." *Journal of Healthcare Informatics*, 12(2), 100-110.
- Patel, P., & Zhao, W. (2022). "Hybrid Model for Cardiovascular Risk Assessment Using Large Language Models and Clinical Datasets." *Journal of Artificial Intelligence in Medicine*, 28(6), 320-328.
- Alvarez, R., et al. (2023). "YOLO-Based Detection of Lung Conditions in X-ray Images." *AI in Healthcare*, 6(4), 257-265.
- Chen, L., et al. (2023). "Speech-to-Text and Machine Learning Integration for Voice-Based Symptom Reporting in Telemedicine." *Proceedings of the International Conference on Health Informatics*, 12(3), 55-63.