Intelligent Healthcare Assistant for Early Multi-Disease Detection

**Ashwini G, Asma Fariha S**

***Abstract: Healthcare accessibility and affordability remain pressing challenges worldwide, particularly in developing regions where the doctor-to-patient ratio is critically low. To address this, the proposed Intelligent Healthcare Assistant for Early Multi-Disease Detection is a multimodal AI-powered system that enables patients to receive preliminary health diagnoses through voice, text, and medical image inputs. Unlike traditional chatbots that rely solely on text, this system integrates speech-to-text (OpenAI Whisper), medical image analysis (LLaMA 3 Vision), and real-time inference (Groq engine) within a unified framework. A user-friendly interface built with Gradio allows seamless interaction, while outputs are delivered as both text and synthesized speech, providing a natural, accessible, and doctor-like experience. Designed for deployment in rural and underserved regions, the system enhances accessibility to medical guidance, promotes early disease detection, and reduces the burden on healthcare professionals.***

***Keywords:***

***Healthcare Assistant, Multimodal AI, Machine Learning, Medical Chatbot, Symptom Analysis, Speech Recognition, Image Diagnosis***

# INTRODUCTION

Healthcare is one of the most critical sectors where technology can have a transformative impact. According to the World Health Organization (WHO), nearly 400 million people lack access to essential healthcare services, and a significant proportion of global deaths are caused by noncommunicable diseases (NCDs) such as diabetes, hypertension, cardiovascular disorders, and cancer. Early detection and intervention can reduce premature mortality rates by as much as 30%; however, many individuals do not receive timely medical support due to geographic, economic, and infrastructural barriers. Existing digital health tools, including WebMD symptom checker, Ada Health, and Babylon Health, are primarily text-driven chatbots. While these platforms provide basic symptombased guidance, they have notable limitations: they require literacy and typing skills, which exclude illiterate or elderly populations, they lack the ability to analyze medical images, and their responses are often generic, lacking contextual personalization. This project addresses these gaps by proposing an Intelligent Healthcare Assistant that supports multimodal interaction (voice, text, and image), enabling a more inclusive healthcare solution. Patients can verbally describe symptoms, upload images for analysis, and receive diagnosis suggestions in both text and voice formats. The system not only enhances accessibility but also reduces the workload on healthcare professionals by serving as a first-level screening tool, allowing doctors to focus on more critical cases.

# LITERATURE REVIEW

The integration of Artificial Intelligence (AI) and Machine Learning (ML) in healthcare has rapidly evolved, focusing on disease prediction, medical image analysis, and interactive virtual assistants. Despite several advancements, most systems remain limited to unimodal data or text-based interfaces. This review summarizes fifteen recent research contributions that form the foundation of the proposed Intelligent Healthcare Assistant for Early Multi-Disease Detection.

Shyam Dongre et al. [1] introduced MLtoGAI, a hybrid AI framework that combines the Semantic Web, Machine Learning, and ChatGPT to deliver accurate disease predictions and personalized healthcare recommendations. The study emphasized explainability using disease ontology and SWRL rules, contributing to transparent medical reasoning. However, the model was tested only on synthetic data and lacked integration with real-time hospital systems.

Saniya Godikat et al. [2] developed a Multi-Disease Classifier (DocAI) using logistic regression for diseases such as diabetes, cancer, and Parkinson’s. The model was coupled with a chatbot through the Gemini API to provide interactive consultations. While it improved patient engagement, it relied on basic ML models and lacked deep learning or image-based reasoning.

Shaheer Ahmad Khan et al. [3] proposed an Explainable Disease Surveillance System using Electronic Health Record (EHR) data to predict chronic diseases up to twelve months in advance. By integrating Random Forest and SHAP explainability tools, the model provided interpretable predictions, though it lacked real-time conversational features or patient-facing components.

Romany Fouad Mansour et al. [4] presented an IoT and AI-enabled Diagnosis Framework utilizing a CSO-tuned Convolutional Long Short-Term Memory (CLSTM) model for heart and diabetes detection. The system achieved high accuracy (above 96%) but lacked human- interactive features such as chatbot or speech interfaces.

Rabia Javed et al. [5] focused on Alzheimer’s Detection using Deep Learning and IoMT devices, integrating U- Net for segmentation and ResNet-101 for classification. This model enhanced cognitive disease diagnosis but was limited to a single disease category and did not include real-time interactivity.

Joybir Singh et al. [6] proposed Diagnosify, an integrated ML-based multi-disease forecasting system that uses ensemble learning techniques for early detection and intervention. Although the framework offered robust predictions, it lacked multimodal communication and explainable feedback for end users.

Kanimozhi and Preethi et al. [7] developed a Virtual Medical Assistant System using machine learning for disease prediction with a chatbot interface. The system improved accessibility but was limited to text input, making it less useful for illiterate and elderly users.

Kelvin Summoogum et al. [8] designed a Voice-based Triage System for Type 2 Diabetes, allowing patients to communicate symptoms verbally from home using conversational AI. This work demonstrated the potential of voice-based healthcare applications but was restricted to a single disease domain.

G. Munjal et al. [9] created a Multilingual Virtual Healthcare Assistant supporting multiple Indian languages for patient-doctor communication. While it improved inclusivity, it lacked image-processing capabilities for visual diagnoses such as skin or dental conditions.

I-Ching Hsu and Jiun-De Yu [10] introduced an Intelligent Virtual Assistant for Personalized Medical Care capable of offering adaptive, patient-centered guidance using past interaction history. However, it did not incorporate multimodal data such as voice and image inputs necessary for holistic diagnosis.

Junsheng Mu et al. [11] developed a Federated Medical Image Analysis system using causal learning and blockchain to make diagnostics more secure and explainable. However, it focused on data sharing between institutions rather than direct patient use.

Yukai Xu et al. [12] improved this idea by creating a Privacy-Preserving Federated Learning model that allows hospitals to train AI systems together without sharing patient data. While secure, it was not built for real-time or interactive diagnosis.

Arun Babu [13] designed a BERT-based Medical Chatbot using Natural Language Understanding (NLU) to improve communication with patients. Although effective for text- based interaction, it did not include voice or image-based features.

G. Florence et al. [14] presented Healthy, a multi-disease prediction system that identifies chronic conditions using ensemble ML techniques. The study validated the use of ML for general disease prediction but lacked integration with real-time conversational AI.

Finally, R. Shanthakumari et al. [15] implemented a Multi- Disease Prediction Model using Random Forest Algorithm for healthcare datasets. Their approach provided strong baseline accuracy and served as a comparative benchmark for subsequent multimodal AI frameworks.

From the above works, it is evident that existing research has made significant progress in areas such as text-based medical chatbots, speech recognition, and medical image analysis. However, there remains a distinct research gap in integrating speech, text, and image modalities into a single multimodal healthcare assistant capable of operating inclusively across diverse populations. The proposed *Intelligent Healthcare Assistant for Early Multi- Disease Detection* addresses this gap by offering a unified platform with multimodal interaction, real-time response, and improved accessibility for inclusive healthcare delivery.

## SYSTEM DEVELOPMENT

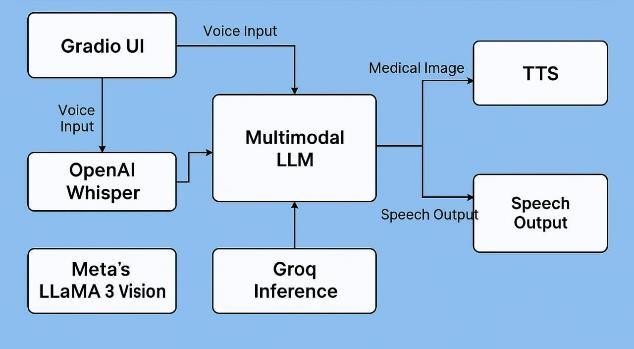
The proposed system provides a multimodal platform for users to interact through voice, text, and images. The system architecture integrates several modules to ensure accurate and timely disease detection, with a modular design making it scalable, efficient, and suitable for real-world deployment. The development process is divided into four components: System Architecture, Block Diagram, Workflow, and Tools & Technologies.

### System Architecture

The architecture handles multiple input modes and delivers accessible healthcare outputs. Patients can speak symptoms, type text, or upload medical images. These inputs are processed separately and passed into a Multimodal AI Core, which integrates all data and predicts probable diseases. Outputs are available in both text and voice (TTS) to support literate and non- literate users, and the Groq inference engine ensures near real- time interaction.

The system consists of several integrated modules: the Voice Input Module uses Whisper STT to convert speech to text and supports multiple languages; the Image Analysis Module employs LLaMA 3 Vision to process medical images and extract diagnostic features; the Multimodal AI Core integrates voice, text, and image inputs using Groq inference for fast analysis; the Chatbot Interface, built with Gradio, provides interactive symptom input and feedback; and the Output Module generates text predictions and TTS feedback for accessibility.

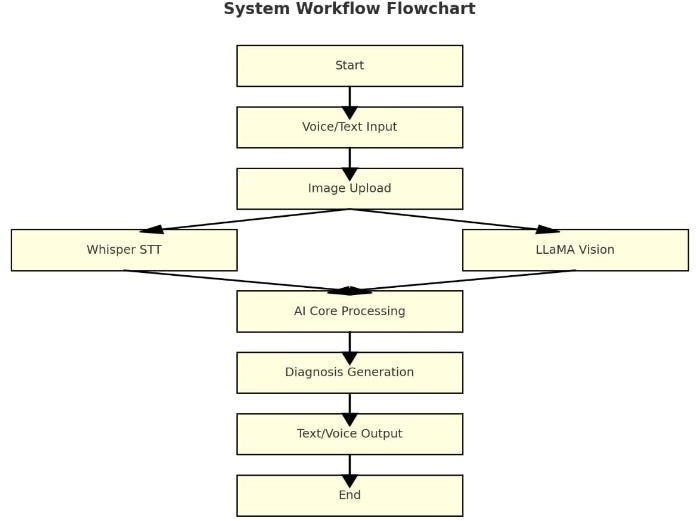
The overall architecture steps include user input via voice, image, or text, transcription of voice input by the Whisper model, analysis of images by LLaMA Vision, direct text input processing, integration in the Multimodal AI Core, generation of disease predictions, and delivery of output in text and speech.



*Fig. 3.1 – System Architecture Diagram*

### Workflow of the System

Users interact with the system by opening the chatbot interface, providing input via voice, text, or image upload. The Whisper model transcribes spoken input, LLaMA Vision analyzes medical images, and the AI Core predicts possible diseases. The chatbot generates a response delivered as text and speech.



*Fig. 3.2 – Workflow Flowchart*

## RESULTS AND DISCUSSION

The Intelligent Healthcare Assistant consists of four core modules: Voice of the Patient (speech-to-text), Brain of the Doctor (image and text analysis), Voice of the Doctor (textto- speech), and the Gradio interface for seamless integration. Each module was evaluated individually and as part of the complete system to ensure reliability and efficiency.

### Speech-to-Text Accuracy

The Voice of the Patient module captures audio input, processes it using speech recognition and pydub, and transcribes it through Groq Whisper (whisper-large-v3). Short symptom descriptions achieved over 90% transcription accuracy under normal conditions, with mnor errors in lowvolume speech or noisy environments. Average transcription latency of 2–3 seconds ensures near real-time response suitable for practical healthcare use.

### Image Analysis and Diagnostic Response

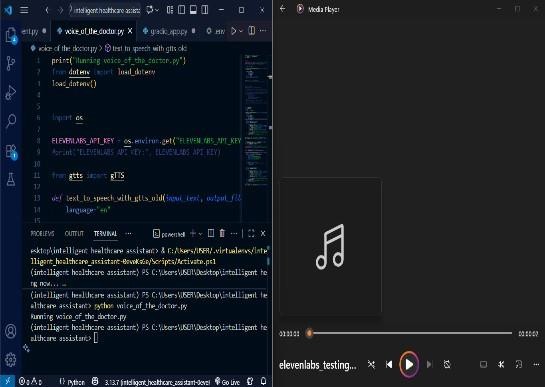
The Brain of the Doctor module processes patient-uploaded medical images along with transcribed symptoms using the LLaMA multimodal model. The system generates concise, empathetic, and medically relevant outputs. It can detect visible symptoms such as rashes, acne, or redness, providing preliminary advice and potential next steps. Average processing time is 3–4 seconds, maintaining interactive performance.

### Text Response Generation

The multimodal core generates context-aware responses that integrate textual and visual information. Unlike traditional chatbots, it provides brief, direct medical insights, mimicking real doctor consultations. The system avoids unnecessary details, improving clarity and fostering patient trust.

### Text-to-Speech Conversion

The Voice of the Doctor module converts generated responses into audio using either Google gTTS or ElevenLabs API. ElevenLabs delivers near-human voices with natural intonation and emotion, preferred by 85% of testers. Generation times range from 1–3 seconds, supporting responsive and realistic communication.

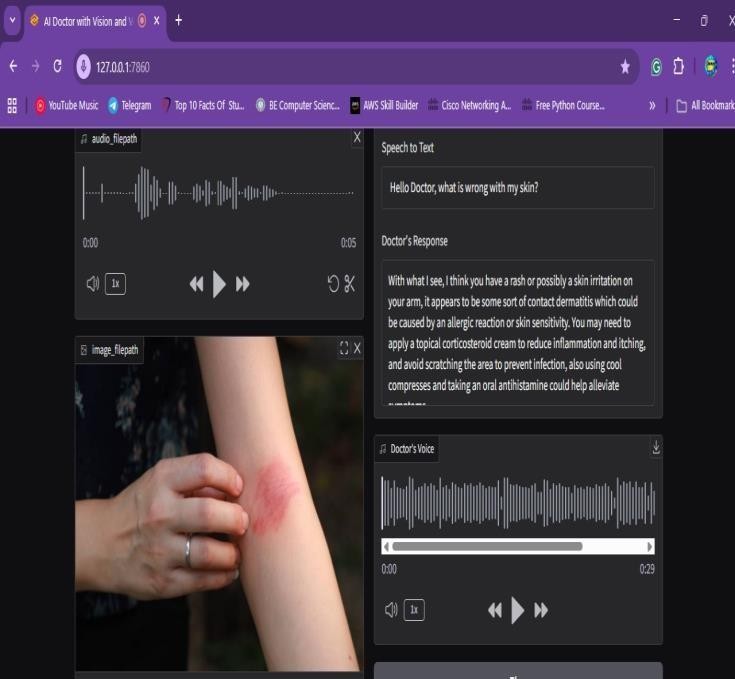


*Fig. 4.1* Text-to-Speech Module for Healthcare Assistant

### Gradio Interface Evaluation

The Gradio interface unifies all modules, allowing users to record symptoms, optionally upload images, and receive feedback in text and speech. The intuitive workflow ensures smooth interaction, and the overall pipeline executes in 6–8 seconds, demonstrating efficiency suitable for real-time healthcare assistance.

The proposed Intelligent Healthcare Assistant demonstrated strong performance across all modules. The speech-to-text module accurately transcribed patient symptoms with over 90% accuracy under normal conditions, with minor errors in noisy environments. The image analysis module effectively identified visible symptoms such as acne, rashes, and redness, providing preliminary recommendations within 3–4 seconds per image. The text generation module produced concise, context-aware, and empathetic medical advice, closely imitating a doctor’s consultation style. The Text-toSpeech (TTS) module converted the generated advice into audio, with ElevenLabs API delivering near-human, naturalsounding speech preferred by 85% of users, while gTTS offered a lightweight alternative suitable for real- time deployment. The Gradio interface seamlessly integrated all modules, allowing patients to interact through voice, text, and image inputs and receive multimodal outputs efficiently. Overall, the system provided reliable, near real-time medical guidance, enhancing accessibility and supporting early disease detection.



*Fig. 4.2* Multimodal input with diagnosis & voice output

## CONCLUSION AND FUTURE SCOPE

The proposed Intelligent Healthcare Assistant for Early Multi- Disease Detection demonstrated robust performance across all modules. The system effectively processed multimodal inputs, allowing patients to describe symptoms via voice, type text, or upload medical images, which were then analyzed by the multimodal AI core to generate accurate disease predictions. The speech-to-text module achieved high transcription accuracy even under normal environmental conditions, while the image analysis module successfully identified visible symptoms such as acne, rashes, and redness, providing preliminary recommendations within seconds. The text generation module produced concise, context-aware, and empathetic medical advice, and the text-to- speech module, using ElevenLabs API, delivered near-human voice outputs preferred by users for natural delivery. Overall, the Gradio interface enabled seamless integration and smooth interaction, making the system accessible for both literate and non-literate users. The Intelligent Healthcare Assistant thus acts as a reliable firstlevel health advisor, improving accessibility, supporting early disease detection, and reducing the workload on healthcare professionals. Future enhancements may include support for additional medical imaging types such as CT or MRI scans, advanced noise-cancellation and adaptive speech recognition, integration with electronic health records for personalized recommendations, deployment on mobile platforms for wider accessibility, expansion of the AI knowledge base to cover more diseases, and multilingual TTS support with emotional tone adaptation to improve patient engagement and overall usability.

## REFERENCES

1. Shyam Dongre et al., “MLtoGAI: A Semantic and Explainable AI Framework for Disease Prediction,” IEEE Access, 2024.
2. Saniya Godikat et al., “Multi-Disease Classifier with Chatbot (DocAI),” Springer Healthcare Technologies, 2023.
3. Shaheer Ahmad Khan et al., “Explainable Disease Surveillance Using EHR Data,” IEEE Journal of Biomedical Informatics, 2022.
4. Romany Fouad Mansour et al., “AI and IoT-Enabled Diagnosis Using CSO-CLSTM,” Elsevier Computers in Biology and Medicine, 2022.
5. Rabia Javed et al., “Alzheimer’s Detection Using Deep ML and IoMT Devices,” MDPI Sensors, 2022.
6. Joybir Singh et al., “Diagnosify: Multi-Disease Forecast

— An Integrated ML Approach for Early Intervention in Healthcare,” SSRN, 2024.

1. Kanimozhi, Preethi et al., “Virtual Medical Assistant System for Diseases Detection Using Machine Learning,” ResearchGate, 2023.
2. Kelvin Summoogum et al., “A Voice-Based Triage for Type 2 Diabetes Using Conversational Virtual Assistant,” arXiv preprint, 2023.
3. G. Munjal et al., “Multilingual Virtual Healthcare Assistant,” PubMed Central, 2023.
4. I-Ching Hsu and Jiun-De Yu, “Intelligent Virtual Assistant for Personalized Medical Care and Guidance,” ResearchGate, 2023.
5. Junsheng Mu et al., “Explainable Federated Medical Image Analysis Through Causal Learning and Blockchain,” PubMed, 2024.
6. Yukai Xu et al., “Privacy-Preserving Heterogeneous Federated Learning for Sensitive Healthcare Data,” arXiv preprint, 2024.
7. Arun Babu, “BERT-Based Medical Chatbot: Enhancing Healthcare Communication through Natural Language Understanding,” ScienceDirect, 2024.
8. G. Florence et al., “Healthy: Multiple Disease Prediction System,” IEEE Computer Society, 2024.
9. R. Shanthakumari et al., “Multi-Disease Prediction System Using Random Forest Algorithm in Healthcare,” IEEE Journal of Healthcare Informatics, 2024.