AI Virtual Health Assistant: Multimodal, Real-Time Symptom Analysis and Risk Prediction

Eric Lin
New York University
el3136@nyu.edu

May 17, 2025

Abstract

Patients face delays and uncertainty when seeking accurate health information. This work introduces an AI Virtual Health Assistant designed to provide instant, personalized health guidance through multimodal input: voice, text, and images. In addition, we integrate a trained XGBoost-based diabetes risk predictor into the assistant, creating a scalable solution for real-time health triage. We demonstrate the system architecture, data processing pipeline, model performance, and future steps for clinical integration.

1 Introduction

Traditional healthcare systems are often overwhelmed, resulting in long wait times and limited access to personalized consultation. Many individuals lack timely guidance for symptom evaluation, leading to poor decision-making and delayed interventions. We address this with a virtual assistant powered by state-of-the-art language and vision models that can interact with users, assess symptoms, and provide voice-based responses.

2 System Architecture

2.1 Frontend

Built using Next.js and Tailwind CSS, the frontend supports dynamic input: text, image uploads, and audio recording. Toast notifications (Sonner) enhance user feedback.

2.2 Backend

The backend is powered by Next.js API Routes:

- Audio is transcribed via Groq Whisper.
- Text and images are processed using Groq's multimodal LLM.
- Voice responses are generated through the Cartesia TTS API.

2.3 Data Flow

1. User submits text, image, or audio.
2. Audio is transcribed if needed.
3. Inputs are forwarded to Groq LLM.
4. The model generates a medical response.
5. The response is spoken using Cartesia and shown on screen.

3 Diabetes Risk Estimation Tool

3.1 Dataset

The CDC Diabetes Health Indicators Dataset includes over 253,000 U.S. survey entries. Features include demographics, health conditions, and behavioral factors. [Ω_datasetΩ_dataset]

3.2 Feature Breakdown

- **Demographics:** Age, Sex, Education, Income
- **Health:** HighBP, HighChol, BMI
- **Behavior:** Smoking, Physical Activity, Alcohol Use
- **Access:** Health Coverage, Doctor Cost Barriers
- **Self-reported:** Physical, Mental, General Health

3.3 Models

We trained and compared three models:

- Logistic Regression
- Random Forest
- XGBoost (selected for deployment)

3.4 Evaluation

Cross-validation (5-fold) was used to evaluate performance. XGBoost achieved the highest accuracy. We present the confusion matrix, classification report, and SHAP analysis.

4 Integration with Gradio

A standalone interface was built using Gradio, allowing users to input lifestyle factors and receive real-time diabetes risk predictions. This integrates seamlessly with the assistant to enrich diagnostic capabilities.

5 Deployment

Hosted on Vercel. Environment variables manage secure API access to Groq and Cartesia services.

6 Conclusion and Future Work

We present a scalable AI health assistant capable of multimodal input handling and intelligent response generation. Future work includes:

- Expanding disease prediction capabilities.
- Finetuning LLMs on medical corpora.
- Enhancing image diagnosis from medical imagery.

References

[1] CDC. Diabetes Health Indicators Dataset.x20; [https://archive.ics.uci.edu/dataset/891/cdc+diabetes+health+indicators](https://archive.ics.uci.edu/