Checkup AI: AI Health Check Partner

Jaffarhussain[1], N S Sandhya[2], Shashank[3], Sindhu Achari[4], Salini Abraham[5]
[1,2,3,4]Students, Department of Computer Science and Engineering, Srinivas Institute of Technology (S.I.T), Mangalore, Karnataka, India
[5]Assistant Professor, Department of Computer Science and Engineering, Srinivas Institute of Technology (S.I.T), Mangalore, Karnataka, India
doi.org/10.64643/IJIRTV12I12-206775-459

***Abstract*—In today's digitally evolving world, gaining access to timely and reliable medical advice is a persistent challenge, particularly for people in rural and underserved regions. The majority of existing digital health tools depend entirely on text-based interaction, which restricts their usefulness for elderly individuals, people with limited typing ability, and anyone dealing with a visible symptom such as a rash, wound, or swelling. This paper presents CheckUp AI, a multimodal virtual health assistant that accepts voice input, uploaded medical photographs, and typed descriptions to generate preliminary health guidance.**

The system integrates automated speech recognition to transcribe spoken symptoms, Base64 image encoding paired with a vision-capable large language model for visual analysis, and a text-to-speech module so users can hear the AI-generated response. A Gradio web interface makes the tool accessible from any standard browser without software installation. Functional and structural testing confirmed that all pipeline components operate correctly and deliver contextually relevant health guidance within seconds. Although Checkup AI is not a substitute for professional medical consultation, it demonstrates a practical, scalable, and inclusive approach to preliminary digital healthcare support.

***Index Terms*—Artificial intelligence, computer vision, Gradio, healthcare chatbot, large language model, multimodal AI, speech recognition, telemedicine.**

I. INTRODUCTION

Obtaining quick and dependable medical guidance remains out of reach for many people worldwide. In rural areas patients often travel considerable distances to reach the nearest clinic, only to wait again for an available doctor. Many who cannot make that journey turn instead to general web searches for self-assessment, despite the well-known limitations of unverified online health information. Digital health assistants have grown in number over recent years, yet almost all of them depend entirely on the user typing out a description of their symptoms an interaction mode that excludes a significant share of potential users.

Elderly individuals, people with limited typing ability, and anyone attempting to describe a visible condition such as a skin rash or a localized swelling find it far more natural to speak their concern aloud or to hold up a photograph. At the same time, traditional rule- based health chatbots struggle with unexpected or complex queries, frequently returning generic, repetitive answers. These shortcomings point to a clear need for an assistant that accepts multiple input types and leverages modern AI to produce nuanced, context-aware responses.

Checkup AI was developed to address this gap. The system combines automated speech recognition, image analysis through a vision-capable large language model, and a text-to-speech output module inside a single unified pipeline, making preliminary health guidance accessible through voice, image, or typed text input. This paper describes the system's motivation, architecture, implementation, testing, and key results.

II. LITERATURE REVIEW

A. AI-Based Healthcare Foundations

The development of artificial intelligence in healthcare has played a crucial role in improving digital medical assistance systems. Early AI-based healthcare models focused on rule-based decision systems, where predefined conditions have been utilized to generate responses.

However, along with the advancement involving machine learning and deep learning, modern systems are capable of analyzing large volumes of medical data and identifying complex patterns. Techniques including natural language processing (NLP) have enabled systems to understand user queries, whereas deep learning models have improved accuracy in symptom analysis. These advancements have laid the foundation for intelligent healthcare assistants capable of providing preliminary medical guidance in real time.

B. Conversational AI and Healthcare Chatbots Conversational

AI has significantly transformed the way users interact with healthcare systems. Early healthcare chatbots relied on keyword-based matching, which often resulted in limited and repetitive responses. With the introduction of advanced NLP models and transformer-based architectures, modern chatbots can understand context and generate more meaningful responses. These systems allow users to describe symptoms in natural language, improving interaction quality. Additionally, AI-powered chatbots are commonly used for symptom checking, appointment scheduling, and patient engagement. However, most current systems still depend largely on text-based communication, limiting their ability to capture detailed and accurate symptom descriptions.

C. Speech recognition and machine learning

Speech recognition technology has enabled more natural and accessible interaction between users and digital systems. By converting spoken input into text, these systems allow users to describe their symptoms without typing, which is particularly beneficial for elderly individuals and users with limited technical skills. In addition, multimodal interaction, which combines text, voice, and image inputs, has gained significant attention in recent research. Integrating multiple input types improves the accuracy and the dependability of AI systems by providing richer contextual information. This method improves user experience and enables more effective analysis of symptoms compared to single-input systems.

D. Medical Image Analysis and AI-Based Diagnosis Support

Medical image analysis has become an important area of research in AI-driven healthcare systems. Deep learning models, especially convolutional neural networks, have demonstrated strong performance in analyzing medical images such as skin conditions, wounds, and infections. These models are capable of detecting patterns and features that might not be easily recognized by non-experts, assisting in early diagnosis and decision-making. Integrating image analysis with conversational AI allows systems. To deliver more precise and context-aware health suggestions. However, challenges such as variability in image quality and limited datasets still affect the overall performance of such systems.
scenes.

E. Intelligent Digital Healthcare Systems

Recent research highlights the increasing adoption of intelligent healthcare systems in telemedicine and remote healthcare applications. These systems aim to provide quick and reliable medical guidance while reducing the burden on healthcare professionals. Key such systems require real-time response generation, high accuracy, user-friendly interfaces, and data security. Cloud-based architectures have further enabled scalable healthcare solutions that can support multiple users simultaneously. The proposed. CheckUp AI system builds upon these advancements by integrating multimodal interaction, AI-based analysis, and an accessible interface to create a practical and efficient virtual healthcare assistant.

III. METHODOLOGY

A. User Input Collection

The application collects two primary inputs simultaneously. A microphone widget captures the user's spoken symptom description, while an image upload widget accepts photographs in JPG, JPEG, or PNG format. An error-handling module validates both inputs before processing begins and informs the user if either is missing or in an unsupported format.

B. Speech-to-Text Conversion

Recorded audio is passed to Python's Speech Recognition library, which processes the signal, suppresses background noise where possible, and returns a plain-text transcription of the spoken description. This text is then combined with the image data and forwarded to the AI analysis stage.

C. Image Preprocessing and Encoding

After validating the uploaded file's format, the image is encoded as a Base64 string. This representation safely embeds binary image data inside a JSON API request without loss or corruption, and is compatible with the Groq API's multimodal input format.

D. AI Model Analysis

The transcribed symptom text and the Base64-encoded image are packaged together and submitted to the Groq API, which runs a large language model capable of processing both text and visual data simultaneously. The model analyses both inputs and generates a response that may include observations about possible conditions, general self-care steps, and a recommendation to seek professional care when the situation appears serious.

E. Response Delivery

The AI response is displayed as formatted text in the interface and simultaneously passed to gTTS, which synthesizes a natural-sounding audio version. Users can listen through built-in playback controls particularly valuable for visually impaired users or those who prefer audio interaction.

IV. SYSTEM DESIGN

A. System Architecture

The Checkup AI architecture spans three layers. The Presentation Layer hosts the Gradio interface and manages all direct user interaction microphone capture, image upload, and output rendering. The Application Layer contains the processing modules: speech recognition, image encoding, AI inference via the Groq API, response formatting, and text-to-speech synthesis. The Data Layer holds the medical knowledge base used during inference and temporary session storage for audio and image files.

All module interactions follow a strict sequential flow. Voice and image inputs arrive at the interface and travel to their respective preprocessing modules. The preprocessed data merges at the AI engine and the resulting response moves through the output modules before reaching the user. An error-handling module monitors every stage and intercepts failures before they propagate downstream.

V. IMPLEMENTATION

The application is written entirely in Python 3.9+. Core libraries include Gradio for the browser-based interface, Speech Recognition for audio transcription, gTTS for speech synthesis, base64 for image encoding, and the Groq client library for API communication. Development was carried out in Visual Studio Code with version control managed through Git and GitHub. The main processing function accepts an audio file and an image file, transcribes the speech, validates and encodes the image, constructs the Groq API payload with both inputs, submits the request, converts the AI response to speech, and returns all four outputs to the Gradio interface: transcribed text, doctor response text, patient voice audio, and synthesized doctor voice audio.

Algorithm Summary:

Step 1: Launch Gradio interface and initialize libraries. Step 2: Capture voice input via microphone. Step 3: Check audio validity; prompt re-record if absent. Step 4: Transcribe audio to text via Speech Recognition. Step 5: Validate uploaded image format (JPG/PNG/JPEG). Step 6: Encode image as Base64 string. Step 7: Combine symptom text and encoded image into API payload. Step 8: Send payload to Groq API; receive AI-generated response. Step 9: Convert response text to audio via gTTS. Step 10: Display text response and enable audio playback.

A. Pseudocode

BEGIN | INITIALIZE gradio_interface, speech_recognition_module, image_processor, AI_model_API, text_to_speech_module| WHILE system_is_running: |audio_input = record_voice() | text = speech_to_text(audio_input) | image = get_uploaded_image()| IF image is not None: | encoded_image = preprocess_and_encode(image) | ELSE: | return "Please upload image"| query = combine(text, encoded_image) | response = AI_model_API.process(query)| audio_response = text_to_speech(response)| display(text, response, audio_response) | END WHILE | END

VI. RESULTS AND DISCUSSION

The CheckUp AI system was implemented and tested to evaluate its ability to assist users in obtaining preliminary medical guidance. The application was tested across multiple scenarios where users described symptoms through voice and shared medical photographs. All pipeline modules speech recognition, image processing, AI analysis, and text-to-speech performed as expected and worked together smoothly.

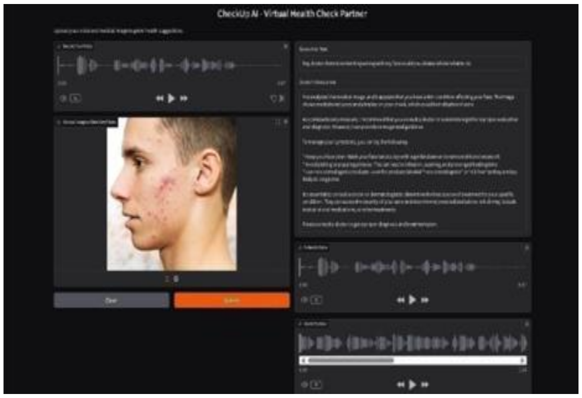

Figure 1: Checkup AI User Interface

Fig. 1 shows the CheckUp AI Gradio interface. It provides three clearly labelled interaction areas: a voice recording section with microphone controls, an image upload section with drag-and-drop support, and an output section containing the speech-to-text panel, the doctor response panel, patient voice playback, and doctor voice playback. The layout cleanly separates inputs from outputs, allowing users to follow the workflow without guidance.

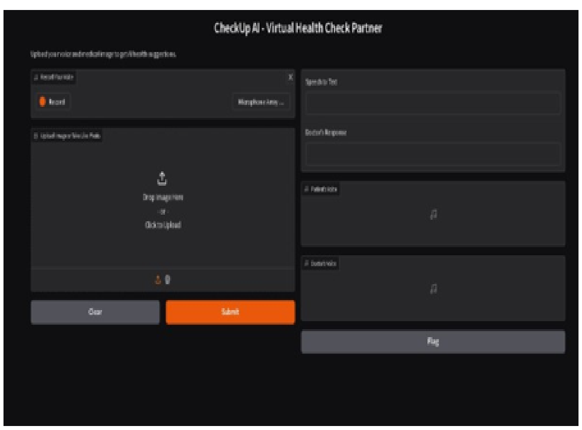

Figure 2: AI Analysis Result

Fig. 2 shows the result of voice input processing. In this test case, the user spoke the phrase "Hey doctor, there is something wrong with my face, could you please tell me what to do." The speech recognition module accurately converted the spoken statement into written text, which was then forwarded to the AI model. The successful transcription confirms that the voice recognition module performs effectively and supports natural interaction without manual typing.

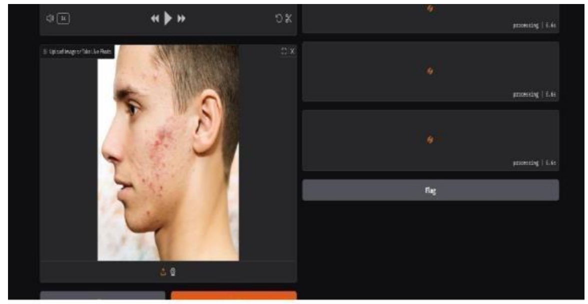

Figure 3: Image Upload and Processing Result

Fig. 3 illustrates the image upload and processing result. During testing, a photograph of facial acne was uploaded alongside the voice input. The image appeared on screen to confirm successful receipt, was validated for format compatibility, and was subsequently encoded in Base64 format for secure transmission to the AI model. The system handled the image upload effectively and confirmed that the image data was transmitted accurately for analysis Fig. 4: AI Generated Medical Response After receiving both the transcribed symptom description and the uploaded image, the Groq-powered model identified a skin condition characterized by multiple red spots and acne-like lesions on the cheek area. The response suggested this was consistent with acne, explained how hair follicles become clogged with oil and dead skin cells leading to inflammation, and provided four practical steps: washing the face twice daily with a gentle cleanser, avoiding picking or popping pimples, using non-comedogenic skincare products, and staying hydrated. The response concluded with a recommendation to consult a dermatologist for a formal evaluation and personalized treatment plan.

Fig. 4 shows the AI-generated doctor's response for the test case.

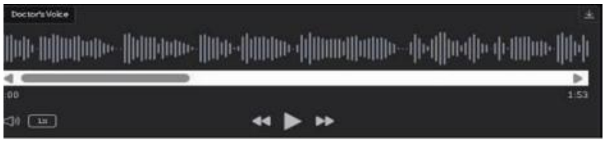

Figure 5: Fig. 5: Voice Output of AI Response

To improve accessibility, the system converts the AI-generated text response into speech using gTTS. Fig. 5 shows the voice output result. The patient voice recording (0:07) and the synthesised doctor voice response (1:53) are both available through the playback controls. Users can listen to the complete medical advice without needing to read the text response, which is especially beneficial for visually impaired users or those who prefer auditory communication.

VII. CONCLUSION

Checkup AI demonstrates that a well-designed multimodal AI pipeline can meaningfully broaden the reach and usefulness of digital health assistance. By accepting voice, image, and text inputs and returning guidance through both text and audio channels, the system serves a more diverse user base than conventional text-only health chatbots. All functional and structural testing confirmed that every component operates reliably and that the integrated system delivers coherent, contextually appropriate health guidance within seconds of input submission. As AI capabilities continue to advance, systems of this kind point toward a future in which quality preliminary health support is available to anyone regardless of location, language, or level of technical familiarity.

ACKNOWLEDGMENT

The authors wish to express their sincere gratitude to the project guide for their valuable guidance, continuous support, and encouragement throughout the design of this project. Their technical suggestions and motivation helped in successfully completing this research work.

The authors also thank the Department of Computer Science and Engineering for providing the necessary facilities and resources required for the development of this project. Special thanks to the faculty members for their academic support and valuable feedback.

The authors would also like to acknowledge the developers of open-source libraries and AI platforms that made this work possible. The availability of modern AI tools and frameworks greatly helped in implementing the CheckUp AI system.

REFERENCES

[1] Topol, "High-Performance Medicine: The Convergence of Human and Artificial Intelligence," Nature Med., vol. 25, pp. 44–56, 2019.

[2] T. Brown et al., "Language Models Are Few-Shot Learners," in Adv. Neural Inf. Process. Syst. (NeurIPS), vol. 33, pp. 1877–1901, 2020.

[3] Vaswani et al., "Attention Is All You Need," in Adv. Neural Inf. Process. Syst. (NeurIPS), pp. 5998–6008, 2017.

[4] Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI Whisper Tech. Rep., 2022.

[1] J. Devlin et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proc. NAACL-HLT, pp. 4171–4186, 2019.

[2] T. Zhao and X. Liu, "Artificial Intelligence in Medical Image Analysis: A Review," IEEE Access, vol. 11, pp. 14523–14540, 2023.

[3] D. Gunning and M. Stefik, "Explainable Artificial Intelligence (XAI)," Sci. Robot., vol. 4, 2019.

[4] World Health Organization, *Ethics and Governance of Artificial Intelligence for Health*. Geneva, Switzerland: World Health Organization, 2021.

[5] Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks," Nature, vol. 542, pp. 115–118, 2017.

[6] OpenAI, "GPT Models Documentation," 2023.

[7] Google, "Google Text-to-Speech API Documentation," 2023.

[8] ElevenLabs, "AI Voice Generation Platform Documentation," 2024.

[9] Gradio Team, "Gradio: Build Machine Learning Web Apps," 2023.

[10] Python Software Foundation, "Python Language Reference," 2023.

[11] S. Rajkomar et al., "Machine Learning in Medicine," N. Engl. J. Med., vol. 380, pp.

1347–1358, 2019.
[12] P. Miotto et al., "Deep Learning for Healthcare: Review, Opportunities and Challenges," Brief. Bioinform., vol. 19, no. 6, pp. 1236–1246, 2018.
[13] J. Jiang et al., "Artificial Intelligence in Healthcare: Past, Present and Future," Stroke Vasc. Neurol., vol. 2, pp. 230–243, 2017.
[14] R. Bibault et al., "Healthcare ex Machina: Are Conversational Agents Ready for Prime Time in Oncology?" Clin. Transl. Radiat. Oncol., vol. 16, pp. 55–59, 2019.
[15] Chen et al., "A Survey on Medical Artificial Intelligence," IEEE Rev. Biomed. Eng., vol. 13, pp. 1–24, 2020.
[16] T. H. Davenport and R. Kalakota, "The Potential for Artificial Intelligence in Healthcare," Future Healthc. J., vol. 6, no. 2, pp. 94–98, 2019.