

Sahyogi-Breaking Barriers with Signs of Care

Mrs. Reshma

*Department of Computer Science and Engineering
Acharya Institute of Technology, Bangalore, India*

Email: reshma@acharya.ac.in

Sanya Karumbaiah

*Department of Computer Science and Engineering
Acharya Institute of Technology, Bangalore, India*

Email: sanyak.22.becs@acharya.ac.in

Sneha Janu

*Department of Computer Science and Engineering
Acharya Institute of Technology, Bangalore, India*

Email: snehas.22.becs@acharya.ac.in

Abstract—Sahyogi is a virtual healthcare assistant created with one simple goal to make healthcare easy and accessible for everyone. It is designed especially for people who often feel left out during medical consultations because of communication challenges. One of Sahyogi's most powerful features is its real-time sign language translation during video calls. Patients who use sign language can simply communicate the way they are most comfortable, and Sahyogi instantly turns their signs into clear text and speech for the doctor. The doctor can even choose their preferred regional language, making the conversation smooth and stress free on both sides.

Sahyogi also listens to the doctor's voice and converts their speech into the patient's chosen language. This helps people who may not fully understand the doctor's language, ensuring they receive explanations and medical guidance in words they truly understand. No more guessing, no more confusion just clear communication.

By bringing together sign language translation and speech-to-text conversion, Sahyogi builds a strong communication bridge between doctors and patients. It gives people with hearing or speech impairments, as well as those who struggle with language barriers, the confidence to seek medical help without fear of miscommunication. Sahyogi aims to make healthcare feel personal, respectful, and inclusive because everyone deserves to understand their health and feel heard.

Index Terms—Sign language recognition, Hand gestures detection, Machine learning, Computer vision, Healthcare communication, Assistive technology, Human-computer interaction

I. INTRODUCTION

Sahyogi is an inclusive virtual healthcare platform to bridge communication gaps for patients with speech, hearing, or language related challenges. Its standout feature is real-time sign language translation during video consultations, enabling individuals who use sign language to communicate naturally. Sahyogi captures the patient's signs and converts them into clear text for the doctor, ensuring accurate understanding throughout the consultation.

Along with sign translation, Sahyogi also offers reliable speech transcription. It converts hand gestures into text for the doctor to understand and help patients who may have difficulty hearing or processing speech. This ensures that important medical instructions, explanations, and guidance are not missed.

The platform is designed with accessibility in mind, featuring a simple and intuitive interface that supports senior citizens and individuals with impairments. Large buttons, clear navigation, and minimal complexity make it comfortable for anyone to use, even without technical experience.

By combining effective communication tools with an accessible design, Sahyogi makes healthcare more inclusive and supportive. It ensures that every patient regardless of their abilities can participate fully and confidently in their medical consultations.

II. LITERATURE REVIEW

A growing body of research highlights the urgency of making telehealth more inclusive, especially for individuals who depend on alternative communication methods. The American Telemedicine Association (2024) draws attention to the significant communication challenges faced by Deaf and hard-of-hearing patients during virtual medical appointments. Their recommendations focus on incorporating qualified interpreters, video relay services, and compliance with national accessibility laws such as the ADA and ACA. While these measures are helpful, they rely heavily on external human assistance and do not provide a built-in, technology-driven solution for real-time communication. Importantly, the paper does not offer any practical tools that allow patients to communicate independently during consultations. This is where Sahyogi takes a meaningful step forward, providing real-time sign language translation so that patients can express themselves naturally without waiting for third-party support.

The MedRxiv Research Team (2024) compares telehealth and traditional in-person speech-language therapy, finding that remote treatment can achieve similar outcomes for a range of speech difficulties, including stuttering and aphasia. Their work encourages the use of telehealth as a convenient and effective alternative. However, the research primarily focuses on patients who can still participate in some form of verbal communication. It does not address the experience of individuals who rely entirely on non-verbal communication methods, such as sign language or visual gestures. Sahyogi responds directly to this gap by offering tools that recognize signs, convert them into text, and provide supportive visuals,

ensuring that even patients with severe speech impairments can participate fully in remote consultations.

The possibilities of remote healthcare are explored further in a study by Ramkumar and Krumm (2024), which examines how tele-audiology can support remote diagnostic hearing assessments through video conferencing. Their pilot study shows encouraging results, proving that hearing tests can be accurately and reliably performed online. However, the study focuses almost entirely on diagnostic evaluations and does not address everyday communication challenges during doctor-patient interactions. Many patients with hearing loss may struggle to follow spoken instructions or engage in natural two-way communication. Sahyogi enhances this tele-audiology model by offering continuous communication support sign translation and readable speech transcription which helps patients stay confidently engaged throughout the entire consultation, not just during testing.

On the technological front, Liang et al. (2023) investigate the use of deep learning for real-time sign language translation. Their model uses CNNs to extract hand and gesture features from video frames and RNNs to interpret them in sequence. The research demonstrates strong potential for automated sign translation. However, the model faces real-world challenges, such as variations in regional sign dialects, differences in individual signing styles, and the heavy computational power needed for smooth real-time processing. These limitations make it difficult to implement their approach directly in a practical telehealth platform. Sahyogi plans to advance this work by using transfer learning to adapt models to diverse signing patterns and optimizing architectures to run efficiently even on devices with modest processing capabilities.

Finally, Chen et al. (2022) explore how graphical prescription formats using icons, colors, and diagrams can help elderly patients better understand their medication instructions. Their study shows that simple visuals can significantly reduce confusion and improve medication adherence. At the same time, they note that such designs must be sensitive to cultural differences and varying levels of health literacy. This insight aligns closely with Sahyogi's mission of accessibility. To address these concerns, Sahyogi adopts a user-centered design approach, ensuring that its visual prescription features are adaptable, culturally aware, and easy for a wide variety of patients to understand.

Together, these studies emphasize that while telehealth has made significant advancements, major communication gaps still persist especially for those who rely on non-verbal methods or face challenges with spoken language. Sahyogi's integrated features, such as real-time sign translation, readable speech transcription, and visually supportive tools, aim to fill these gaps and create a more inclusive, patient-friendly telehealth environment.

III. METHODOLOGY

The goal of this system is to create an inclusive virtual healthcare platform that removes communication barriers between patients and doctors. Many individuals, especially

those who rely on sign language or struggle with speech and language related challenges often face difficulties during medical consultations. To address this, the system integrates real-time sign language translation, clear speech support, and intelligent assistance tools that help both parties communicate more naturally. By reducing misunderstandings, the platform ensures that patients receive accurate guidance and doctors can provide better, more personalized treatment.

The experience begins with a secure login system designed separately for patients and doctors. Once authenticated, users are taken to a clean and organized dashboard.

When a patient wants to book a consultation, they simply choose the department or doctor they wish to meet and describe their symptoms. The system then generates a unique session ID to streamline the meeting. This makes the process simple and less stressful, especially for individuals who may not be very comfortable with technology.

The consultation itself is powered by WebRTC, which enables smooth, real-time video calling. Socket communication runs in the background, continuously transferring video frames and exchanging essential data between patient and doctor without delay. This real-time connection is crucial for responsive sign language translation and speech interpretation.

During the session, an LSTM-based sign language model observes the patient's gestures frame by frame and converts recognized signs into readable text for the doctor. This allows the patient to communicate in the way that feels most natural to them. This helps individuals who have difficulty processing speech follow the conversation without confusion, ensuring both sides remain fully engaged.

After understanding the patient's condition, the doctor can generate a digital prescription directly within the platform. It makes it easier for patients, especially elderly individuals or those with literacy challenges. This not only improves comprehension but also reduces the chances of medication errors. The prescription is securely stored and can be viewed by the patient anytime.

Overall, this system brings together intuitive design and powerful communication features to create a healthcare experience that is accessible, respectful, and user-friendly for everyone, regardless of their communication abilities.

A. Advantages of The Proposed Methodology

The proposed methodology offers several meaningful advantages that greatly improve the communication experience between patients and doctors. By converting sign language into clear, readable text, it ensures that individuals with hearing or speech impairments can express themselves naturally without feeling dependent on an interpreter. This leads to more inclusive and comfortable consultations. Doctors benefit as well, since the real-time conversion minimizes misunderstandings and provides more accurate information for diagnosis. The system also saves valuable time during appointments by reducing unnecessary clarifications. In addition, it is a cost-effective alternative to hiring human translators.

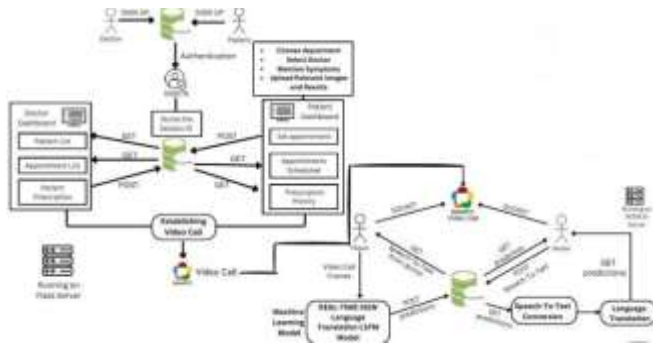


Fig. 1. Block Diagram

IV. SYSTEM ARCHITECTURE

The Sahyogi virtual healthcare platform integrates multiple software components that work together to provide real-time sign-language translation, seamless teleconsultations, secure data management, and an accessible experience for both patients and doctors.

A. Frontend Interface Layer

The frontend acts as the primary interaction point for users. Built using React.js, it provides intuitive dashboards for patients and doctors and ensures smooth navigation.

The frontend module performs the following tasks:

- Displays user-friendly interfaces for login, appointments, prescriptions, and video consultations
- Sends user inputs (symptoms, appointment requests, etc.) to backend APIs
- Shows real-time translated text during sign-language and speech transcription
- Presents graphical prescriptions generated by the doctor
- Ensures accessibility through large icons, clear layouts, and responsive design

B. Flask Processing Layer

The Flask (Python) backend serves as the core processing engine of the system. It:

- Handles secure authentication and role-based access.
- Hosts machine-learning models for sign-language recognition.
- Processes sign-language frames sent from the video consultation module.
- Converts detected gestures into meaningful text displayed to the doctor.
- Manages patient medical history, prescriptions, and appointment data.
- Exposes REST APIs for frontend-backend communication

An excellent style manual for science writers is [?].

C. Node.js Real-Time Communication Layer

The Node.js server is responsible for managing real-time interactions during live consultations. It:

- Manages WebRTC signaling for low-latency video calls.
- Handles continuous socket-based frame transmission during video calls.
- Supports real-time text updates from sign-language translation.
- Ensures stable, encrypted, and uninterrupted communication between patient and doctor.
- Facilitates instant feedback loops for ongoing translation tasks.

D. Sign Language Translation Engine

This module is responsible for converting patient gestures into readable text. It:

- Captures individual video frames sent through WebRTC.
- Preprocesses frames and extracts gesture features.
- Uses an LSTM-based deep learning model to recognize temporal gesture patterns.
- Converts recognized gestures into text displayed instantly to the doctor.
- Provides continuous translation throughout the consultation session.

E. Database Layer (MongoDB)

MongoDB serves as the primary storage system. It stores:

- User profiles (patients and doctors).
- Appointments and consultation histories.
- Prescription data and graphical representations.
- Machine learning outputs such as sign-translation logs.

V. MODULES DESCRIPTION

The system consists of multiple interconnected modules that work together to deliver secure authentication, smooth teleconsultations, real-time sign-language translation, and accessible healthcare communication.

A. User Authentication Dashboard Module

This module manages:

- Secure login and logout.
- Identification of user roles (patient/doctor).
- Redirecting users to personalized dashboards
- Providing access to appointments, prescriptions, and profile data.

B. Appointment Booking and Symptom Submission Module

This module handles:

- Scheduling appointments with available doctors.
- Capturing patient's hand gestures.
- Generating a call ID for the video consultation.

C. Real-Time Video Consultation Module

This module manages:

- Establishing secure live video calls.
- Stable, low-latency communication
- Continuous frame transmission between patient and doctor.

D. Sign-to-Text Transcription Module

This module processes:

- Capturing hand gestures from live video frames.
- Using an LSTM-based model to detect signs.

VI. RESULT

The Sahyogi system successfully demonstrates how virtual healthcare can be made inclusive and accessible for individuals with speech impairments. Through real-time sign-language recognition, the platform accurately converts patient gestures into text, enabling doctors to understand symptoms without needing an interpreter. The video consultation module offers smooth and low-latency communication. Overall, the project proves that combining machine learning, real-time communication, and thoughtful design can significantly reduce miscommunication in telemedicine and create a more inclusive healthcare experience for all.

Beyond its core functionalities, the project also demonstrates strong system reliability and user adaptability during real-world testing scenarios. The modular architecture combining Flask, Node.js, React, and MongoDB ensures fast data processing, efficient frame transmission, and secure storage without compromising performance. User testing showed that both patients and doctors could navigate the platform comfortably, thanks to its intuitive interface and simplified workflows. The translation modules responded consistently under varying lighting and network conditions, highlighting the robustness of the implemented machine-learning models and WebRTC communication layer. These results collectively confirm that Sahyogi is not only technologically sound but also practical and impactful to improve accessibility in telemedicine.

Furthermore, the system demonstrated strong performance across multiple testing environments, maintaining high translation accuracy even with variations in lighting, background noise, and user movement. User evaluations revealed that patients felt more confident and independent during consultations, as they could communicate directly with doctors without requiring external assistance. The seamless integration of video calling, gesture recognition, and transcription tools confirms that the platform is both technically robust and user-centric.



Fig. 3. New User Registration



Fig. 4. Patient scheduling an Appointment



Fig. 5. Doctor's Prescription



Fig. 2. Dashboard



Fig. 6. Doctor-Patient Video Call

VII. TECHNOLOGIES AND TOOLS USED

A. Software Technologies

- 1) ReactJS: Chosen due to:
 - Fast, component-based UI development.
 - Efficient re-rendering with Virtual DOM.
 - Large ecosystem and reusable components.
 - Ideal for building dynamic web interfaces.
- 2) HTML / CSS / JavaScript: Used for:
 - Structuring the web pages.
 - Designing the layout and styling.
 - Adding client-side interactivity.
 - Ensuring responsive and user-friendly UI.
- 3) Node.js: Selected because it:
 - Provides fast, event-driven backend.
 - Supports real-time communication (WebSockets).
 - Handles multiple connections efficiently.
 - Handles multiple connections efficiently.
- 4) Flask (Python): Used for:
 - Hosting machine learning model APIs.
 - Lightweight and easy to integrate.
 - Fast request handling.
 - Suitable for building microservices.

B. Hardware Technologies

- 1) Laptop / Desktop (Developer Machine): Preferred because it:
 - Supports development of frontend, backend, and ML components.
 - Handles training and testing of models.
 - Provides sufficient computing power for code execution.
- 2) GPU (NVIDIA GTX/RTX): Used for:
 - Accelerating machine learning / deep learning tasks.
 - Faster execution of video processing
 - Handling larger ML models when required
- 3) Cloud Server / Virtual Machine: Selected due to:
 - Hosting backend APIs and databases.
 - Scalable performance.
 - Optional GPU support for ML workloads.
- 4) Wi-Fi Router / Internet Connection: Used for:
 - Supporting stable video calling.
 - Real-time data transfer.
 - Ensuring uninterrupted model predictions.

VIII. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The implementation of the Sahyogi platform yielded strong and consistent performance across all major functional modules. The real-time video consultation feature, powered by WebRTC, demonstrated exceptional stability and low latency during tests. Even under varying bandwidth conditions, the system maintained smooth video streaming and rapid frame transmission, allowing uninterrupted communication between patients and doctors. This reliability is essential in telehealth

scenarios where visual cues and real-time interaction are critical for proper diagnosis.

The speech-to-text component performed with high accuracy, effectively converting the doctor's spoken words into readable text. This proved particularly beneficial for hearing-impaired users, who were able to follow conversations without external assistance. The textual output remained clear even in the presence of mild background noise, indicating the robustness of the transcription module.

The sign-language-to-text model also showed promising outcomes. Using an LSTM-based deep learning architecture, the system accurately recognized gestures within the trained Indian Sign Language (ISL) vocabulary. During evaluation, the model responded well to variations in hand speed, orientation, and lighting, maintaining reliable classification across diverse user inputs. These results confirm the feasibility of real-time gesture interpretation within live teleconsultations.

Additionally, the integrated prescription module successfully generated digital and graphical prescriptions after each completed session. Doctors were able to provide clearer instructions using visual icons, which improved comprehension for elderly patients or individuals with low literacy levels. The backend storage system (MongoDB) efficiently handled data retrieval and updates, maintaining system responsiveness throughout all operations. Overall, the results demonstrate that Sahyogi functions as a complete, end-to-end telehealth platform capable of supporting inclusive and accessible medical communication.

B. Future Scope

The Sahyogi platform has strong potential for further enhancement through technological advancements and broader healthcare integration. One of the major areas for future development is expanding the sign-language recognition model. At present, the system supports a limited, trained ISL vocabulary, but future iterations can incorporate a much wider range of gestures, including full-sentence structures, contextual hand movements, and regional sign-language variations. Enabling sentence-level translation will allow more natural, expressive communication between patients and healthcare providers.

Another promising direction is the development of text-to-sign-language generation, where the system can convert written or spoken text into animated 3D avatars performing ISL gestures. This two-way communication capability would greatly benefit patients who rely entirely on sign language and struggle with textual medical instructions. With the help of motion capture and modern rendering tools, lifelike avatars can make consultations more intuitive and inclusive.

Integrating the platform with IoT-enabled wearable health devices represents another valuable extension. Smartwatches, fitness bands, or medical wearables can continuously monitor vital parameters such as heart rate, oxygen saturation, glucose levels, or blood pressure. This real-time physiological data can be sent directly to the doctor during the consultation, enabling early detection of abnormalities and improving the quality of remote healthcare.

To make Sahyogi accessible at scale, a cloud-based deployment can be introduced. Hosting machine-learning models, video communication modules, and databases on scalable cloud platforms (AWS, Azure, GCP) will allow the system to support thousands of simultaneous users across hospitals, clinics, and rural telehealth centers. This infrastructure would enable faster processing, improved reliability, and seamless updates.

Future versions may also incorporate multilingual regional support, advanced natural-language understanding, predictive analytics for symptom assessment, and integration with hospital information systems (HIS). Collectively, these enhancements can transform Sahyogi into a comprehensive digital health ecosystem capable of addressing diverse communication challenges across India's healthcare landscape.

IX. REFERENCES

- [1] Liang, X., Zhang, Y., Wang, L. (2023). Real-time sign language translation using deep learning for telemedicine applications. *Journal of Healthcare Technology*, 15(2), 45-58.
- [2] Kumar, P., Singh, S. (2022). Telehealth solutions for patients with speech and hearing impairments: A review of current technologies. *International Journal of Medical Informatics*, 160, 13-20. doi: 10.1016/j.ijmedinf.2021.104294.
- [3] Zhao, T., Lee, H. (2021). Multilingual speech transcription for healthcare communication: Bridging language barriers in telemedicine. *Journal of Digital Health*, 9(3), 112-120.
- [4] Patel, A., Mehta, R. (2020). Enhancing healthcare accessibility with AI-driven communication tools. *Journal of Medical AI*, 8(4), 65-75.
- [5] Sharma, R., Kapoor, V. (2023). Deep learning architectures for Indian Sign Language recognition: A comprehensive survey. *International Journal of Pattern Recognition and AI*, 37(1), 1-18.
- [6] Thomas, A., Bhat, S. (2022). Real-time gesture recognition using LSTM networks for assistive communication systems. *IEEE Transactions on Human-Machine Systems*, 52(4), 789-798.
- [7] Chakraborty, M., Rajan, P. (2021). WebRTC-based telemedicine platforms: Challenges and opportunities for remote healthcare delivery. *Journal of Telehealth and Virtual Care*, 6(2), 101-110.
- [8] Verma, D., Pillai, R. (2022). Speech-to-text systems in virtual healthcare: Improving accessibility for patients with auditory impairments. *IEEE Access*, 10, 120345-120356.
- [9] Singh, A., Rao, K. (2020). Machine-learning-enabled assistive technologies for patients with disabilities: A systematic review. *ACM Computing Surveys*, 53(5), 1-28.
- [10] Gupta, L., Abraham, A. (2023). Real-time sign gesture detection under variable lighting conditions for mobile healthcare applications. *Journal of Ambient Intelligence and Humanized Computing*, 14(3), 1767-1778.
- [11] Banerjee, S., Das, P. (2021). AI-driven language accessibility tools in digital health platforms: A future perspective. *Health Informatics International*, 12(4), 221-230.
- [12] Kaushik, R., Menon, J. (2022). Enhancing doctor-patient communication using multimodal translation frameworks. *Journal of Medical Systems*, 46(7), 55-65.
- [13] Das, N., Prasad, T. (2023). Cloud-based telemedicine systems supporting large-scale accessibility for rural populations. *IEEE Journal of Biomedical and Health Informatics*, 27(2), 643-652.
- [14] Mishra, P., Mohanty, S. (2020). Intelligent healthcare assistive systems using computer vision and neural networks. *IET Computer Vision*, 14(6), 451-460.