

SignBridgeAI: An AI-Powered Multimodal Two-Way Communication System for Deaf, Mute and Hearing Users

Shreya Ghodake¹, Sadiya Gavandi², Kshiti Gambhir³, Gauri Bangale⁴, Prof. Nikhil Deshpande⁵

¹Department Of Information Technology, Sinhgad College of Engineering, Pune- 41

²Department Of Information Technology, Sinhgad College of Engineering, Pune- 41

³Department Of Information Technology, Sinhgad College of Engineering, Pune- 41

⁴Department Of Information Technology, Sinhgad College of Engineering, Pune- 41

⁵Department Of Information Technology, Sinhgad College of Engineering, Pune- 41

Email: shreyaghodake.scoe.it@gmail.com

Abstract - Communication barriers between deaf, mute, and hearing individuals continue to limit social inclusion and accessibility in everyday interactions. While existing solutions provide partial support, many are either hardware-dependent, costly, or lack real-time bidirectional capabilities.

This paper presents SignBridgeAI, an AI-powered, lightweight, and browser-based multimodal communication system designed to enable seamless interaction using Indian Sign Language (ISL). The system integrates real-time sign-to-text recognition, text-to-speech conversion, and text-to-sign (ISL fingerspelling) using web technologies. Unlike traditional approaches, it operates entirely within a web environment without requiring specialized hardware.

The proposed system uses MediaPipe for hand tracking, a trained deep learning model for gesture classification, and browser-based APIs for speech synthesis. Additionally, deployment on the Internet Computer enables decentralized, serverless, and always-available access.

Experimental evaluation demonstrates efficient real-time performance with high usability and low latency, making the system suitable for practical communication scenarios. The proposed solution contributes toward scalable, affordable, and inclusive assistive technology.

Key Words: Indian Sign Language (ISL), Gesture Recognition, Deep Learning, Text-to-Speech, Sign-to-Text, Human-Computer Interaction, MediaPipe, Accessibility, Decentralized Deployment, Assistive Technology

1. INTRODUCTION

Communication is a fundamental human necessity that enables social interaction, learning, and participation in daily life. However, individuals who are deaf or mute face significant communication barriers due to the lack of widespread understanding of sign language among hearing individuals. Indian Sign Language (ISL), though widely used within the deaf community, is not commonly understood by the general population, resulting in limited accessibility and social inclusion.

Traditional solutions rely on human interpreters, which are often unavailable or expensive. Recent advancements in

artificial intelligence, computer vision, and natural language processing have enabled the development of automated systems for sign language recognition. Previous work such as SignBridgeAI demonstrated the feasibility of multimodal communication systems integrating gesture recognition and speech processing.

However, these systems still face limitations such as:

- Dependence on external APIs and backend servers
- High computational requirements
- Limited accessibility due to installation or hardware needs

To address these challenges, this paper presents SignBridgeAI, a fully web-based, decentralized communication system that operates directly in the browser without requiring specialized hardware. The system integrates gesture recognition, speech synthesis, and ISL visualization into a single platform, enabling seamless interaction between deaf, mute, and hearing users.

2. LITERATURE SURVEY

[1] This research presents the Saudi Deaf Companion System (SDCS), a comprehensive platform for two-way communication between deaf and hearing individuals using Saudi Sign Language (SSL). The system integrates a sign recognition module based on a custom 3D Graph Convolutional Network (3DGCN), a speech recognition and synthesis module (SRSM), and an avatar module (AM) for visual animation of SSL. With the introduction of the largest SSL dataset (KSU-SSL, comprising 293 signs from 10 domains), the SDCS achieves high recognition accuracy, scalable text-to-sign conversion, and speech synthesis. The platform enables deaf individuals to participate more fully in education, healthcare, and societal activities and addresses challenges of interpreter scarcity by using automated avatars, thus promoting social integration and autonomy for the deaf community.

[2] This paper proposes SignSpeak, a real-time system translating sign language to speech via a wearable glove embedded with flex sensors and spatial tracking. Sensor data is processed by machine learning models to recognize and classify gestures, which are then rendered into synthesized

speech. The design is affordable, portable, and supports user-defined gestures for personalization. Validation studies show SignSpeak achieves a recognition accuracy of 94.7% across 29 gesture classes and an end-to-end response time of 112 ms, supporting practical real-time communication. The paper provides a comparative analysis of previous translation systems and demonstrates that SignSpeak overcomes cost and adaptability limitations seen in existing solutions, positioning it as an accessible and scalable communication aid.

[3] This study introduces SignTone, an AI-based system designed to break communication barriers for the deaf using advanced computer vision, deep neural networks, and voice processing. The proposed system is composed of a Gesture Interpretation Component (GIC), using deep convolutional networks for real-time sign language video processing, and a Verbal Communication Component (VCC) for bidirectional speech and text conversion. The framework also features a dynamic virtual assistant providing real-time visual translations. Experimental results demonstrate high gesture recognition accuracy (92%), robust speech-to-text performance (89%), and strong virtual assistant animation reliability (94%). The project highlights advancements in multimodal learning, spatiotemporal gesture analysis, and user-centric accessibility, indicating significant promise for making communication truly inclusive for deaf individuals.

[4] This paper presents a machine learning-based approach for sign language conversion to text and speech. The system leverages convolutional neural networks for real-time gesture recognition, complemented by natural language processing and text-to-speech synthesis to produce expressive spoken output. Designed to handle variations across sign languages, this approach emphasizes adaptability and scalability for inclusive communication.

3.PROPOSED SYSTEM

The proposed system, SignBridgeAI, is a multimodal communication platform designed to enable seamless interaction between users through three core functionalities:

1. Sign-to-Text Conversion

Hand gestures are captured using a webcam and processed using MediaPipe. Extracted landmarks are passed to a trained deep learning model, which predicts the corresponding ISL character and displays it as text.

2. Text-to-Speech Conversion

Users can input text, which is converted into speech using the browser's built-in SpeechSynthesis API. Adjustable speech rate enhances usability

3. Text-to-Sign (ISL Fingerspelling)

Typed text is converted into ISL alphabet representations using pre-mapped visual assets (A–Z), enabling visual communication.

3.METHODOLOGY

A. Data Collection: Gesture data is captured using a webcam under different lighting conditions and backgrounds.

B. Preprocessing: Frames are processed using MediaPipe to extract 21 hand landmark points.

C. Feature Extraction: Landmark coordinates are normalized to ensure consistency.

D. Model Training: A Convolutional Neural Network (CNN) is trained to classify ISL alphabets.

E. Real-Time Inference: The trained model predicts gestures in real time from live video feed.

F. Confidence Mechanism: To reduce noise:

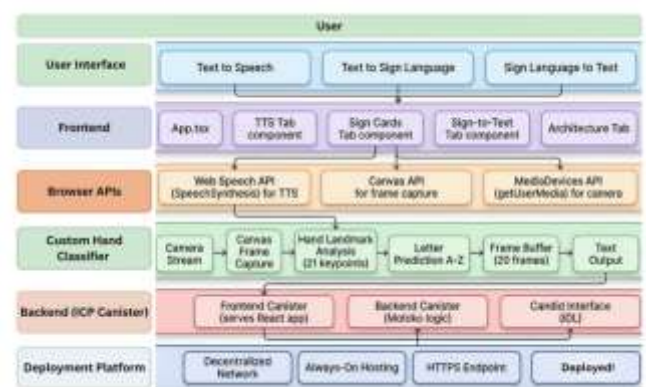
- Predictions collected over 600 ms
- If $\geq 70\%$ match $\rightarrow$ accepted

G. Text-to-Speech: Speech synthesis is performed using browser APIs.

H. Text-to-Sign Mapping: Each character is mapped to ISL visual representation.

4.SYSTEM ARCHITECTURE

The proposed system follows a modular, layered, and decentralized architecture designed to support real-time multimodal communication using web technologies and AI-based gesture recognition.



1. High-Level Functional Modules:

At the top layer, the system is divided into three core functionalities:

- Text-to-Speech (TTS): Converts user-input text into audible speech using browser-native APIs.

- Text-to-Sign Language: Transforms text into Indian Sign Language (ISL) fingerspelling representations using visual sign cards.

- Sign Language to Text: Recognizes hand gestures via camera input and converts them into readable text output.

These modules ensure bidirectional communication between deaf/mute and hearing users.

2. Frontend Application Layer

The frontend is implemented using a React-based architecture (App.tsx) and is structured into multiple tab components:

- TTS Tab Component: Handles text input and triggers speech synthesis.
- Sign Cards Tab Component: Displays ISL alphabet representations for typed text.
- Sign-to-Text Tab Component: Captures real-time video and performs gesture recognition.
- Architecture Tab: Provides system visualization and documentation.

This modular UI design ensures clean navigation, scalability, and accessibility.

3. Web API Integration Layer

The system leverages browser-native APIs for efficient real-time processing:

- Web Speech API (SpeechSynthesis): Used for text-to-speech conversion without external dependencies.
- Canvas API: Captures video frames for processing gesture input.
- MediaDevices API (getUserMedia): Accesses the webcam for real-time video streaming.

These APIs eliminate the need for external services, making the system lightweight and fast.

4. Gesture Processing Pipeline

The core AI pipeline for sign recognition operates as follows:

- Camera Stream Acquisition:
Live video is captured from the user's webcam.
- Frame Capture using Canvas:
Frames are extracted at regular intervals for processing.
- Hand Landmark Detection:
MediaPipe extracts 21 keypoints representing hand structure.
- Feature Processing & Prediction:
The extracted landmarks are passed to a trained model to classify gestures into ISL alphabets (A–Z).
- Temporal Buffering:
A frame buffer of 20 frames (~600 ms) is maintained.
- Confidence Filtering:
Predictions are validated only if $\geq 70\%$ of frames agree,

ensuring robustness and reducing noise.

- Text Output Generation:

Final recognized characters are displayed as text.

This pipeline ensures high accuracy, low latency, and real-time performance.

5. Decentralized Backend Architecture (ICP)

The system is deployed on the Internet Computer Protocol (ICP) using canister-based architecture:

- Frontend Canister:
Hosts and serves the React web application.
- Backend Canister (MotokoLogic):
Handles application logic, data processing, and coordination.
- Candid Interface (IDL):
Defines communication between frontend and backend.

This architecture provides:

- Decentralized execution
- Serverless deployment
- High availability and scalability

6. Deployment Layer

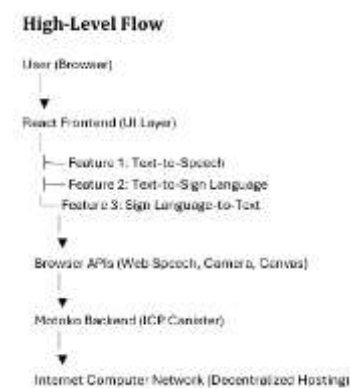
At the lowest layer, the system operates on a decentralized infrastructure offering:

- Always-On Hosting
- Secure HTTPS Endpoints
- Distributed Network Availability

This ensures the system is accessible globally without dependency on centralized servers.

5. SYSTEM DESIGN

1.Design Overview



2. Module design

Layer 1 — User Interface (Frontend)

Technology: React + TypeScript + Tailwind CSS + Vite

Component	Purpose
App.jsx	Root component; renders heading and tab container
Tab: Text to Speech	Input box + speak button + speed slider + word count
Tab: Text to Sign	Input box → renders ISL fingerspelling cards (A–Z)
Tab: Sign to Text	Camera feed → hand classifier → text output

Data flow within frontend:

- State managed locally using React useState and useEffect hooks
- No external state library (Redux, Zustand) -- kept simple
- Tabs switch views without page reload (single-page app)

Layer 2 — Browser APIs (Native, No External Server)

Technology: Web Speech API, MediaDevices API, Canvas API

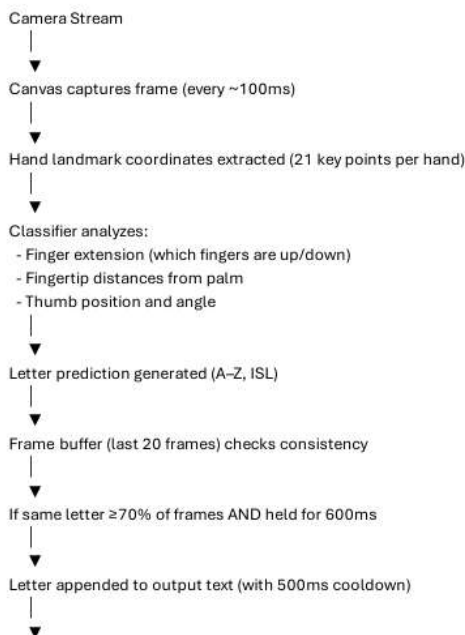
API	Feature	What it does
Web Speech API (SpeechSynthesis)	Text to Speech	Converts typed text to audio using the device's built-in voice engine
MediaDevices API (getUserMedia)	Sign to Text	Requests camera permission and streams live video
Canvas API	Sign to Text	Captures individual video frames from the camera stream for analysis

Layer 3 — Custom Hand Classifier

Technology: Custom JavaScript/TypeScript

This is the core intelligence of the Sign-to-Text feature.

Flow:



Confidence bar shown to user (gray → amber → green)

Layer 4 — Backend (ICP Canister)

Technology: Motoko

The backend is a Motoko canister deployed on the Internet Computer. It handles:

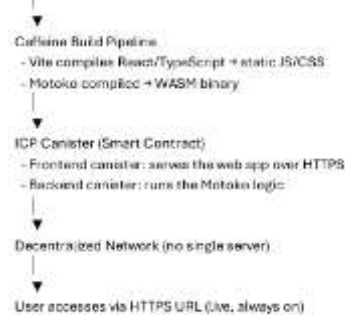
Function	Purpose
Serving the frontend	The canister serves the compiled React app as static assets
Backend API	Exposes any server-side functions callable from the frontend
State persistence	Any data that needs to persist across sessions lives here

The frontend communicates with the backend via **Candid interface** (ICP's IDL), auto-generated as `backend.d.ts`.

Layer 5 — Deployment Platform

Technology: Caffeine + Internet Computer Protocol (ICP)

Source Code (React + Motoko)



3. Data Flow

Feature 1: Text to Speech

- User types text
- React state updates
 - User adjusts speed slider (0.5x–2x)
 - User clicks "Speak"
 - Web Speech API (`SpeechSynthesisUtterance`) called
 - Browser plays audio through device speakers
 - Word count displayed in real time

Feature 2: Text to Sign Language

- User types text
- React state updates
 - Each character mapped to ISL fingerspelling card
 - Cards rendered as visual components (A–Z images/representations)
 - User sees fingerspelling sequence on screen

Feature 3: Sign Language to Text

- User clicks "Start Camera"
- `MediaDevices` API requests camera permission
 - Live video stream rendered in `<video>` element
 - Canvas captures frame every 100ms
 - Hand landmark coordinates analyzed by custom classifier
 - Letter prediction made (A–Z, ISL)
 - Frame buffer tracks last 20 predictions
 - If ≥70% consistent + held 600ms → letter confirmed
 - Letter appended to output text
 - User can Copy / Backspace / Add Space

6.IMPLEMENTATION RESULTS

1.Web Results



Img 1: Sign To Text

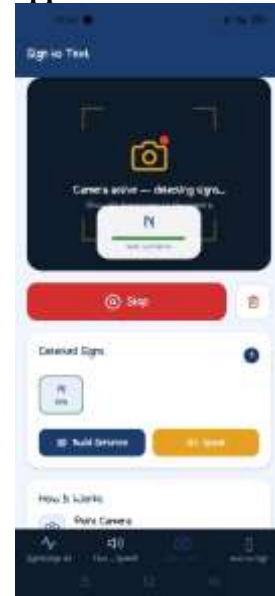


Img 2: Text To Sign



Img 3 : Text To Speech

2. App Results



Img 1: Sign To Text



+



Img 2: Text To Sign and Text To speech

7.CONCLUSION AND FUTURE SCOPE

This paper presented SignBridge AI, an innovative AI-powered multimodal platform that bridges communication gaps among deaf, mute, blind, and hearing individuals via real-time sign language recognition, speech-to-text conversion, and animated avatar translations. The system achieves high accuracy, low latency, and broad accessibility using standard consumer devices. Pilot testing indicates improved communication efficacy and user acceptance.

Future work will focus on expanding the sign language vocabulary to include regional and contextual variations, incorporating advanced natural language processing for nuanced interpretation. Enhancements in avatar realism and emotion expression will be pursued to enrich interactions. Additionally, integrating tactile or haptic feedback could broaden accessibility for deaf-blind users. Further scalability improvements aim to support larger user bases and offline capabilities. Overall, SignBridge AI is poised to make a substantial societal impact by enabling inclusive and effective communication for differently-abled communities.

8. REFERENCES

- [1]. M. Faisal, M. Alsulaiman, M. Mekhtiche, M. Bencherif, M. Algabri, T. Alrayes, G. Muhammad, H. Mathkour, W. Abdul, Y. Alohal, M. Al-Hammadi, H. Altaheri, and T. Alfakih, "Enabling two-way communication of deaf using Saudi sign language," *IEEE Access*, vol. 12, pp. 1–18, 2023, doi: 10.1109/ACCESS.2023.3337514.
- [2]. I. Marin, V. Babalain, and S. Slutu, "Real-Time Sign Language-to-Speech Translation System With AI- Powered Wearable Technology," *IEEE Access*, vol. 13, pp. 151652–151663, Sept. 2025, doi: 10.1109/ACCESS.2025.3602794.
- [3]. A. M. Ravishankkar, B. Balavishwa, S. Chandru, M.S. Dharinendra, and M. Sanjay, "AI-powered Signtone deaf companion system for inclusive communication between deaf and hearing individuals," *Journal of Emerging Trends and Novel Research*, vol. 3, no. 4, Apr. 2025, pp. a153–a160.
- [4]. A. M. Anitha, P. Annanya, R. C. Venkata Shritha, K. Hargun, and M. H. S. Mangala, "Sign Language Conversion to Text and Speech Using Machine Learning," *International Journal of Innovative Research in Technology*, vol. 10, no. 12, pp. 2846-2851, May 2024.
- [5]. H. Kim, I. Lee, S. Panda, S. Hajra, B. Jeong, J. Seo, K. R. Kaja, M. A. Belal, V. Vivekananthan, and H. J. Kim, "Smart gloves-based triboelectric nanogenerator for sign language detection," *Micro and Nano Systems Letters*, vol. 13, no. 11, 2025, doi: 10.1186/s40486-025-00231-7.