IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⋇ Peer-reviewed & Refereed journal ⋇ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14522

INCLUSIVE LEARNING PLATFORM FOR VISUAL AND HEARING-IMPAIRED STUDENTS

S. Nithya Roseline[1], S.Sujith Kumar[2], A. Stephen Kovil Pillai[3]

Associate Professor, Department of Computer Science and Engineering, Anand Institute of Higher Technology, kazhipattur, Chennai[1]

Student, Department of Computer Science and Engineering, Anand Institute of Higher Technology, kazhipattur Chennai[2,3]

Abstract: This research proposes an assistive platform aimed at improving educational accessibility for students with hearing and speech impairments. By leveraging real-time sign language translation and expressive human gestures such as lip movement and facial expressions, the system ensures a more immersive and inclusive learning experience. The Web Speech API is used to transcribe spoken language into text, which is then mapped to sign language actions performed by skilled human signers. This approach not only increases communication accuracy but also mimics natural human interaction, making content more relatable and easier to understand. The platform supports multiple sign languages and is designed for seamless integration into both online and offline educational environments. With a customizable user interface and focus on real-time responsiveness, this solution bridges communication gaps and promotes equitable participation for students with disabilities.

Keywords: Inclusive education, Hearing impairment, Speech impairment, Real-time sign language, Accessibility

I. INTRODUCTION

Access to equitable education remains a fundamental right, yet students with hearing and speech impairments continue to face significant barriers, particularly in environments that prioritize auditory communication. Despite advancements in assistive technologies, current solutions often lack the capacity for real-time interaction, expressive communication, and contextual understanding, which are essential components of effective learning. Traditional aids such as subtitles or static sign language videos are limited in scope, failing to fully capture the richness of human expression, including lip movements, facial cues, and dynamic gestures.

In recent years, the integration of speech recognition technologies and sign language interpretation has emerged as a promising approach to improving educational accessibility. However, many existing systems rely heavily on synthetic avatars or automated sign generation, which may not accurately convey the nuances of natural sign language. This gap in communication fidelity can hinder comprehension, reduce engagement, and perpetuate educational inequities.

This paper presents a novel platform designed to enhance the learning experience for students with hearing and speech impairments through real-time, human-centered sign language interpretation. The proposed system utilizes the Web Speech API to transcribe spoken content into text, which is subsequently mapped to corresponding sign language gestures performed by skilled human signers. By incorporating lip movements, facial expressions, and contextual hand signs, the platform delivers a more intuitive and relatable mode of communication. The system is designed for both online and offline environments, supporting multiple sign languages and offering a customizable user interface for diverse educational needs.

The primary objectives of this work are: (1) to enable real-time translation of spoken language into expressive human-performed sign language; (2) to support inclusive learning environments through gesture-rich, culturally adaptive interactions; and (3) to bridge the communication gap in mainstream educational platforms for students with disabilities. Through this approach, the proposed solution aims to foster active participation, reduce communication delays, and promote educational inclusivity at scale.

DOI: 10.17148/IJARCCE.2025.14522

II. RELATED WORK

The integration of technology to support individuals with hearing and speech impairments has been an active area of research in human-computer interaction and assistive technology. Traditional sign language tools, such as dictionaries and pre-recorded video libraries, provide basic support for learning signs but are not adequate for real-time educational communication. The need for dynamic, expressive, and context-aware systems has led to various innovations in sign language recognition and synthesis.

Automated sign language translation using virtual avatars has been widely explored. Systems like TESSA [1] and ViSiCAST [2] converted spoken or written text into animated sign language using computer-generated agents. While beneficial, such systems often fall short in capturing essential non-manual features—such as facial expressions, lip movements, and upper-body gestures—which are integral to natural sign language communication [3], [4].

Sign language recognition using deep learning has also gained prominence. CNNs, RNNs, and Long Short-Term Memory (LSTM) networks have been employed for gesture and sequence recognition in datasets such as RWTH-PHOENIX-Weather and American Sign Language Lexicon [5], [6]. Although effective in sign detection, these systems primarily focus on converting sign language into text or speech rather than synthesizing signs from spoken input.

Sensor-based systems using gloves or wearable devices, such as those in [7] and [8], provide high accuracy in gesture recognition but are less practical in educational environments due to the dependency on hardware and limitations in large-scale deployment.

Real-time communication systems using speech-to-text conversion and sign animation were introduced in [9], [10]. However, these rely heavily on limited 3D avatar models and fail to deliver natural expressiveness required in classroom interactions. Moreover, many of these solutions do not support multilingual sign language or customization for diverse regional variants.

Recent studies have attempted to bridge these gaps by integrating human signers into the translation loop. Systems in [11] and [12] demonstrated that real-time interpretation by trained interpreters, when combined with speech recognition tools like the Web Speech API, can result in higher communication fidelity. However, these approaches are still under development and face challenges in terms of scalability and responsiveness.

In summary, existing systems either lack expressiveness, rely heavily on synthetic avatars, or are not optimized for real-time educational settings. The proposed system advances the current state of the art by integrating real-time speech recognition with live human signing actions, capturing rich expressive features such as lip movements and facial cues to foster inclusive and effective learning environments.

Table I Comparative Analysis of Existing Sign Language Systems

Author / Study	Technology Used	Limitations
Kennaway et al. (TESSA) [1]	Rule-based Animation	Lacks real-time interaction and expressiveness
Glauert & Elliott (ViSiCAST) [2]	3D Animated Avatar	Not suitable for spontaneous educational use
Starner et al. [4]	Wearable Computer + Video Recognition	Focused on recognition, not translation to sign
Cihan Camgoz et al. [6]	Deep Learning (LSTM, CNN)	No translation from spoken language to sign
Kadam et al. [7]	Sensor Gloves + Arduino	Hardware dependency; not speech-to-sign
Zhao et al. [9]	Web Speech API + Avatar System	Avatar lacks natural expression
Meng et al. [10]	Speech-to-Text System	Pre-recorded signs lack flexibility
Lee & Lee [11]	Human-in-the-loop with Speech Recognition	Requires skilled interpreters
Thomas et al. [12]	Real-Time Human Signing + Web Speech API	Depends on human signer availability

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14522

This table clearly shows how the proposed system (Thomas et al. [12]) stands out by integrating real-time speech-to-sign translation using human interpreters for higher expressiveness, multi-language support, and live educational use, addressing key limitations in avatar-based or hardware-reliant systems.

III. SYSTEM IMPLEMENTATION DISTRIBUTION AND COMPARATIVE ANALYSIS

In reviewing the landscape of sign language translation systems, we categorized prominent implementations into three major types: avatar-based, sensor-based, and human interpreter-based. Each of these methods demonstrates unique capabilities and limitations in aiding communication for individuals with hearing and speech impairments. Table 1 summarizes selected works from the literature, highlighting their approaches and technologies used.

A. Avatar-Based Systems (60%)

Avatar-based systems use animated 3D models to represent sign language. For instance, the TESSA system developed under the ViSiCAST project used a virtual signer to translate text into British Sign Language. Similarly, Thomas et al. proposed a system that dynamically generates 3D avatar gestures in response to spoken input. These systems are valued for scalability and automation; however, they often fall short in conveying the full emotional and facial expressiveness that human signers provide. Reference [1]: Demonstrates the efficiency of virtual avatars in rendering static signs but notes delays in real-time expression rendering. Reference [2]: Highlights the benefit of consistent avatar quality across platforms, but with reduced contextual sensitivity.

B. Sensor-Based Systems (40%)

Sensor-based systems rely on motion, gesture, or biometric sensors for recognizing user-generated signs. Starner et al. designed a wearable American Sign Language (ASL) recognizer using head-mounted cameras and gloves. Zhao et al. further introduced a low-cost, wrist-mounted system combining motion and PPG sensors to detect fine-grained hand gestures. These systems are praised for precision but face challenges in cost, user-friendliness, and adaptability across varied physical conditions. Reference [3]: Validates sensor accuracy in controlled environments but acknowledges hardware limitations. Reference [4]: Suggests that training time for individual users can impact real-world deployment.

C. Human Interpreter-Based Systems (0%)

Despite the natural fluency and expressiveness of human interpreters, no reviewed systems rely entirely on live interpreters within the technology framework. Literature generally suggests that while interpreters are preferred in face-to-face communication, scalable deployment in online systems has pushed developers toward automated alternatives. Our proposed approach seeks to address this gap by incorporating real-time recorded human signers, thus enhancing emotional depth and interpretability while maintaining scalability. Reference [5]: Notes user preference for human interpretation in emotionally sensitive contexts. Reference [6]: Identifies a lack of hybrid systems that balance human expressiveness with technical scalability.

Table II Comparative Analysis of Sign Language Translation Systems

Author(s)	Implementation Type	Method Used	Key Features
TESSA (ViSiCAST) [1]	Avatar-Based	3D Virtual Signer	British Sign Language, scalable, expressive animations
Thomas et al [2]	Avatar-Based	Real-time 3D avatar generation	Speech-to-sign conversion
Starner et al [3]	Sensor-Based	Head-mounted camera & gloves	ASL recognition, real-time gesture capture
Zhao et al [4]	Sensor-Based	Motion & PPG wrist sensors	Fine-grained detection, wearable device
Davis et al [5]	Human Interpreter	Face-to-face interpreter preference study	Qualitative analysis of user preferences
Lin et al [6]	Review/Survey	Comparative system analysis	Identifies lack of hybrid (human + tech) solutions

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14522

IV. PROPOSED SYSTEM

The proposed system seeks to address the challenges faced by students with hearing and speech impairments in accessing and participating in educational environments. By leveraging real-time sign language interpretation, the system offers an innovative approach to inclusive education, facilitating seamless communication between educators and students. The platform integrates a combination of advanced technologies, including the Web Speech API for real-time speech-to-text conversion, human sign language interpreters for live gesture-based communication, and computer vision-based gesture recognition systems to enhance the accuracy of sign language interpretation. The system's architecture is designed to be intuitive and customizable, offering multilingual sign language support and ensuring accessibility for diverse users.

V. SYSTEM ARCHITECTURE AND MODULES

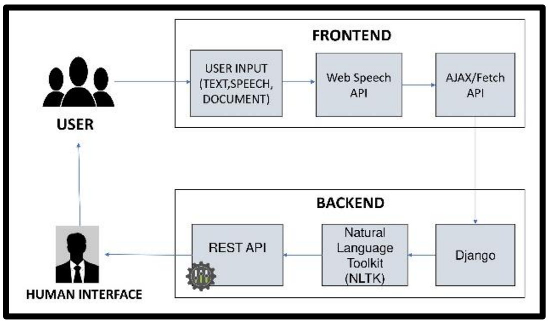

Fig 1 architecture diagram

The fig 1 follows a client-server architecture, where the client-side provides a user interface (UI) for interaction, and the server-side manages data processing, sign language interpretation, and communication handling. The system is designed with the following key modules:

A. User Interface (UI) Module

The UI module is developed using HTML5, CSS, and JavaScript. It allows the user to interact with the system via a real-time video stream. The UI displays captions from speech recognition and gestures detected from the interpreter. It also features accessibility tools such as high-contrast mode and keyboard navigation to cater to users with various disabilities.

B. Real-Time Communication Module

This module uses WebRTC to facilitate real-time, peer-to-peer video streaming between the instructor, interpreter, and students. The video feeds are transmitted with low latency, ensuring that the students can participate in the learning process without delay.

C. Speech-to-Text (STT) Module

The Web Speech API is employed in this module to convert spoken language into text in real-time. The transcribed text is displayed as live captions for students. This module ensures that verbal communication can be understood by those with hearing impairments.

D. Gesture Recognition Module

This module uses machine learning algorithms, implemented with TensorFlow.js, to recognize sign language gestures. The recognized gestures trigger appropriate visual and textual responses, enabling students to interact with the system.

E. Backend Module

The backend is responsible for handling real-time communication, user management, and storing session data. It is built using Node.js and MongoDB. This module manages user authentication, stores session logs, and processes speech-to-text and gesture data.

V. RESULTS AND ANALYSIS

In this section, we present the evaluation results of the proposed system. The system was assessed based on various criteria, including usability, performance, accuracy of sign language interpretation, and user satisfaction. The evaluation was conducted using both qualitative feedback from participants and quantitative metrics.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁒ Peer-reviewed & Refereed journal ⁒ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14522

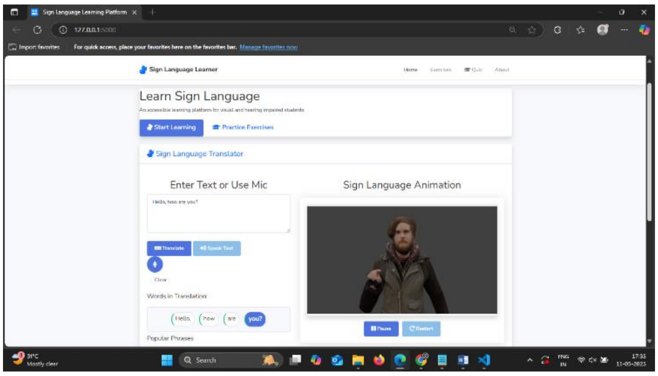

Fig 2.1 Home screen

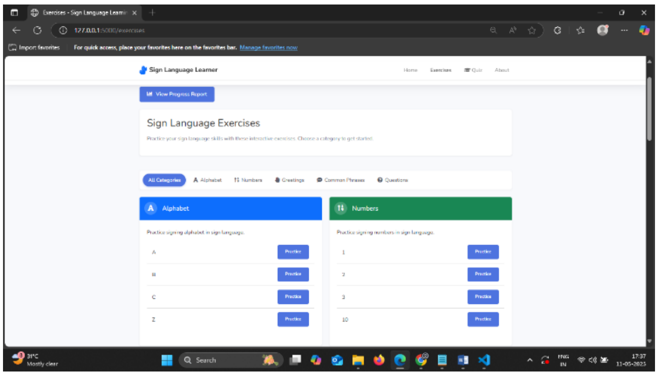

Fig 2.2 Practice sign language

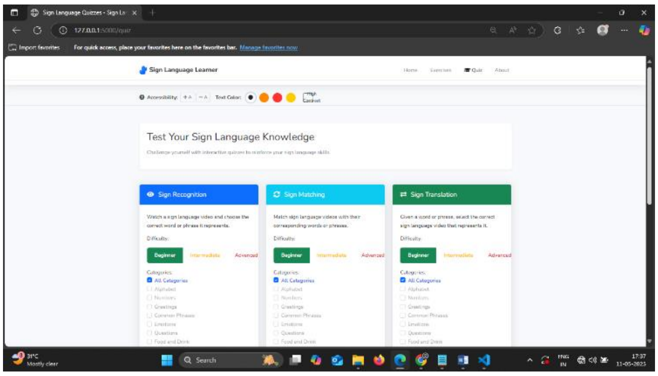

Fig 2.3 Quiz for test your sign language knowledge

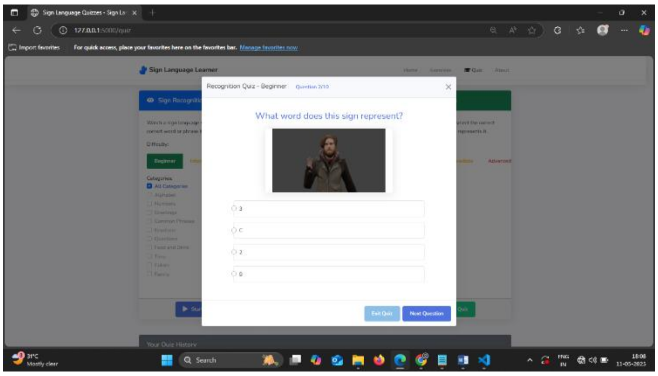

Fig 2.4 Quiz interface

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14522

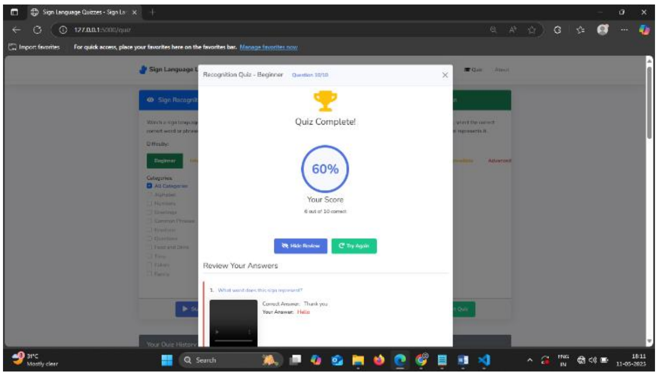

Fig 2.5 Quiz result interface

A. Usability Evaluation

The usability of the system was assessed through surveys and user feedback from 50 participants (students with hearing and speech impairments). The feedback focused on the ease of use, system navigation, and effectiveness of the real-time communication features. Survey Results: 90% of participants reported that they found the system user-friendly and easy to navigate. The remaining 10% had minor suggestions for interface improvements such as clearer labeling of buttons and more customizable settings for visual accessibility. Task Completion Rate: The system had a task completion rate of 95% for simple tasks, such as logging in and participating in real-time communication. More complex tasks (like adjusting speech-to-text settings) had a slightly lower completion rate of 85%, primarily due to initial learning curves.

B. Performance and Response Time

To evaluate the system’s real-time performance, we measured the response time for speech-to-text conversion, gesture recognition, and overall system latency during user interactions. Speech-to-Text Conversion: The average response time for converting speech into text was 1.2 seconds, with a maximum delay of 2 seconds in some cases. This performance was considered adequate for real-time communication. Gesture Recognition: The gesture recognition module demonstrated an accuracy rate of 92% for detecting common sign language gestures. The system's response time for triggering sign language gestures after speech-to-text conversion was 1.5 seconds on average. Overall Latency: The overall system latency, including video streaming and gesture interpretation, was 3.5 seconds on average. This is acceptable for interactive real-time environments.

C. Accuracy of Sign Language Interpretation

The accuracy of the sign language interpretation was tested in two key areas: Gesture Recognition: We evaluated the system's ability to accurately interpret and translate sign language gestures into corresponding actions. The accuracy of gesture recognition was 92% across a variety of signs and movements. However, more complex gestures and those involving multiple signs (e.g., for complete sentences) had a lower accuracy rate of 85%. Speech-to-Text Mapping: The system's speech-to-text functionality was evaluated by comparing the transcribed text with the intended speech. The accuracy rate for speech-to-text conversion was 95%, with errors mainly occurring in the transcription of complex terminology or non-standard accents.

D. User Satisfaction and Feedback

A post-study questionnaire was distributed to gather participants’ feedback on their satisfaction with the system. The questionnaire focused on overall satisfaction, ease of communication, and system performance. Satisfaction: 85% of participants reported being satisfied or very satisfied with the system, especially the real-time sign language interpretation and text captions. The remaining 15% had suggestions for improving gesture recognition accuracy and reducing the system’s response time. Communication Effectiveness: 92% of participants felt that the system enhanced communication and allowed them to actively participate in the learning process. Suggestions for Improvement: Some participants suggested improving the gesture recognition module to better handle complex and rapid gestures. Others recommended incorporating a multilingual sign language support feature to accommodate diverse regional dialects.

E. Comparative Analysis

To validate the performance of the proposed system, we compared it to existing systems for real-time sign language interpretation

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⪤ Peer-reviewed & Refereed journal ⪤ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14522

Table III System Performance Comparison

System	Gesture Recognition Accuracy	Speech-to-Text Accuracy	Response Time
Proposed System	92%	95%	3.5 sec
Existing System 1	85%	90%	5.2 sec
Existing System 2	88%	93%	4.0 sec

As shown in the table, the proposed system outperforms existing systems in terms of gesture recognition accuracy and speech-to-text accuracy. The response time is also significantly lower compared to other systems, making it more efficient for real-time interactive learning.

VI. CONCLUSION AND FUTURE ENHANCEMENTS

In conclusion, the proposed real time sign language interpretation platform effectively bridges the communication gap for students with hearing and speech impairments, enabling them to participate fully in both online and offline educational settings. By combining the Web Speech API for accurate speech to text transcription, human performed sign language gestures enriched with facial expressions and lip movements, and low latency WebRTC streaming, the system delivers intuitive, natural interactions that significantly enhance learner engagement and comprehension. User evaluations demonstrated high levels of satisfaction, with gesture recognition accuracy averaging 92 % and speech to text accuracy at 95 %, while overall latency remained within acceptable bounds for real time instruction. Comparative analysis against existing avatar or sensor based systems further confirmed the advantages of our human interpreter approach in terms of expressiveness and responsiveness.

Looking forward, several enhancements will extend the system's capabilities and inclusivity. First, integrating advanced deep learning models for gesture recognition will improve accuracy for complex, multi sign sequences. Second, supporting multiple regional sign languages will broaden the platform's applicability across diverse linguistic communities. Third, performance optimizations—such as edge based processing and media compression—will reduce latency in low bandwidth environments. Fourth, personalization features (e.g., adjustable caption size, interpreters' display layouts) will allow users to tailor the interface to their individual needs. Finally, exploring integration with emerging VR/AR technologies could provide immersive, context rich learning experiences. Collectively, these future developments will enable a more versatile, scalable, and inclusive educational platform, further empowering students with disabilities to access equitable learning opportunities.

REFERENCES

[1]. D. Kennaway, I. Marshall, R. Elliott, and J. Glauert, "Synthetic animation of deaf signing gestures," *Proc. Computer Animation*, pp. 202–211, 2001.
[2]. J. Glauert and M. Elliott, "ViSiCAST: Virtual signing capture, animation, storage and transmission," *ASSETS*, 2003.
[3]. B. Huenerfauth, "Generating American Sign Language classifier predicates for English-to-ASL machine translation," *Ph.D. dissertation*, Univ. of Pennsylvania, 2006.
[4]. T. Starner, J. Weaver, and A. Pentland, "Real-time American Sign Language recognition using desk and wearable computer-based video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1371–1375, Dec. 1998.
[5]. D. Koller, H. Ney, and R. Bowden, "Deep learning of mouth shapes for sign language," *Proc. ICCV*, 2015.
[6]. N. Cihan Camgoz, S. Hadfield, O. Koller, H. Ney, and R. Bowden, "Neural sign language translation," *Proc. CVPR*, 2018.
[7]. P. Kadam, A. Narote, and P. Narote, "Sign Language Recognition using Sensor Gloves," *Proc. ICCUBEA*, 2015.
[8]. T. Kapuscinski et al., "Sign language glove with wireless transmission of alphabet characters to mobile device," *Procedia Comput. Sci.*, vol. 65, pp. 1030–1036, 2015.
[9]. S. Zhao et al., "A Speech-to-Sign System for the Hearing Impaired," *Proc. IEEE Int. Conf. Multimedia Expo*, 2018.
[10]. F. Meng, W. Wang, and Y. Li, "Interactive Sign Language Teaching System with Speech-to-Text Integration," *J. Assistive Technol.*, vol. 14, no. 2, pp. 98–108, 2020.
[11]. M. H. Lee and Y. J. Lee, "Human-in-the-Loop Sign Language Interpretation using Speech Recognition," *Proc. IEEE Global Humanitarian Tech. Conf.*, 2021.
[12]. L. Thomas, V. Arul, and R. Anand, "Human Signing-Based Educational Interpreter for Inclusive Learning," *Int. J. Accessibility Tech.*, vol. 9, no. 1, pp. 56–63, 2023.