

AI-Based Indian Sign Language Translator

Aditya Kumawat^{1,*}, Chitranshi Agarwal¹, Akshita Jain¹, Chirag Jain¹, Tanisha Jain¹

Abstract

Indian sign language (ISL) is the primary form of communication for a large number of deaf and hard-of-hearing people within the Indian subcontinent. However, their interactions with non-signers remain limited, mainly due to the absence of tools that could translate sign language in real time. The present study aims to develop an ISL translation system powered by AI that can automatically convert sign gestures into readable text or spoken output. The proposed system will analyze video input with the help of computer vision and deep learning techniques, estimate hand motion along with facial cues, and finally recognize ISL gestures with a high accuracy level. It deals with three major aspects: increasing recognition performance across users and environments, removing delays so as to enable real-time communication, and designing an easy-to-operate interface that could be used inside classrooms, hospitals, workplaces, and public service centers. The system is envisioned to bridge the communication gap between signed and spoken/written languages as a means to support inclusive communication and advance the accessibility of the Indian deaf community in various digital and physical spaces.

Keywords: Indian sign language, deep learning, computer vision, gesture recognition, real-time translation, accessibility

INTRODUCTION

One of the largest problems that deaf individuals encounter in the Republic of India every day is communication. Indian Sign Language (ISL) is an established, comprehensive language that deaf and hard-of-hearing individuals use instinctively and effortlessly [1]. However, the thing is, very few individuals outside the deaf community know how to speak ISL. Very few hearing individuals have learned the language. Thus, if the deaf want to communicate, for example, in the classroom, hospital, office, or with the government, things become rather more complicated [2]. Usually, the deaf are forced to ask someone else for help in relaying the information, or they have no choice but to write things down.

This is essentially about creating a computer system that will help solve problems. It involves creating a device that is capable of monitoring the video of a person interpreting through ISL and transmitting that particular message through text or language [3, 4]. It does not need an interpreter, and there is no need for it to be put down. Communication is required.

*Author for Correspondence

Aditya Kumawat

E-mail: adityakumawat.ai27@jecrc.ac.in

¹Student, Department of Artificial Intelligence and Data Science (AIDS), JECRC Foundation (Jaipur Engineering College and Research Centre), Sitapura, Jaipur, Rajasthan, India.

Received Date: February 03, 2026

Accepted Date: February 20, 2025

Published Date: March 10, 2026

Citation: Aditya Kumawat, Chitranshi Agarwal, Akshita Jain, Chirag Jain, Tanisha Jain. AI-Based Indian Sign Language Translator. Journal of Image Processing & Pattern Recognition Progress. 2026; 13(2): 23–31p.

However, this is only one part of an even bigger agenda, which is being targeted at this point. This is an aspect of technology because it facilitates deaf and hearing-impaired persons in Indian society to get an equally effective mode of communication. This also makes a commitment to ensure that education, healthcare facilities, job opportunities, and governing systems are also available to them. We are also working on the

agenda of making it possible so that it would be able to process continuous signs. In addition, this would be a full-fledged conversation, which would take place in real time so that translation would start immediately when one person starts to sign details from his or her side [4, 5].

BACKGROUNDS

Indian Sign Language Overview

ISL is the primary medium of communication among deaf and hearing-impaired persons in India. It is a full-fledged natural language with its own grammar, syntax, and vocabulary, which uses hand gestures, facial expressions, and body language [1]. ISL differs from all other natural languages because it is both visual and gesture-based and can communicate complex information and knowledge sans sound [3]. There are also different versions or dialects of ISL in different states and among various groups of people, which makes it more complex for communication by humans and machines [2].

The creation of automated translation systems requires in-depth knowledge of the linguistics and semantics of ISL so that accurate interpretation or translation of certain gestures can take place.

Existing Sign Language Translation Systems

During the last decade, many approaches have been proposed by researchers and developers to fill this communication gap between the sign language-speaking community and other communities. Their approaches were categorized as follows:

Audio to Sign Translation Systems

This system translates audio into text, and sign language gestures are created using animated avatars, robotic hands, or videos. This system is useful in public services, where announcements or instructions are translated into sign language [3].

Sign-to-Text/Audio Systems

In these systems, computer vision algorithms are used to interpret gestures, facial expressions, and body language. These gestures are identified using machine learning algorithms and translated into text or speech. These systems allow real-time interaction between ISL and non-ISL users [4, 5].

Hybrid Bidirectional Systems

Some contemporary models have incorporated the use of both for seamless two-way communication between users of ISL and non-sign-language users. In this case, deep learning models, such as convolutional neural networks for gesture recognition and recurrent neural networks for modeling sequences, have been employed for increased accuracy and adaptability [4].

Although these models perform well in a controlled setup, they face challenges in terms of scalability, ability to handle different ISL dialects, and performance in a practical setting. Moreover, the existing models are largely dependent on central servers and, to a large extent, pose potential risks to issues of data privacy, ownership, and sustainability [2, 6].

Limitations of Existing Systems

Despite the impressive technological advances achieved during the last few decades, the following challenges still exist in the current frameworks for sign language translation:

Limited Vocabulary and Regional Adaptability

Current systems tend to represent a limited set of signs, and regional variations in ISL are ignored [1, 2].

Real-Time Processing Requirements

The processing of visual gestures and the generation of real-time translation cannot be efficiently performed by various systems because of the intensive nature of processing, causing translation to be delayed or inaccurate [4, 5].

Centralized Data Dependence

In most cases, when existing systems are considered, their dependence on centralized data in servers for hand and model parameter information poses problems of privacy, data manipulation, and control [2]. Unoptimized Web 3.0 web-based decentralized technology uses: Web 3.0 is not optimized for decentralized web-based technology, such as blockchain technology and cloud storage systems [7]. In the event that Web 3.0 is not used, it is considered non-transparent, insecure, and cannot be scaled. In this context, a multifaceted approach is required, one in which AI translation and Web 3.0 technology are integrated to ensure that ISL communication is safe and efficient [6, 8]. ISL communication would not only be made possible but would also enable deaf individuals to become more productive members of their societal, academic, and work settings.

Problem Statement

Audio to ISL is a visual and gesture-based language used for communication across the country. It includes extensive gestures of hands, face, and body postures to convey the message. ISL helps direct communication and provides deaf people with opportunities to associate with each other for emotional expression, sharing ideas, and other daily dealings.

Description

The requirement of audio to Indian Sign Language conversion tool emerges because of the communication gap and lack of understanding between deaf people who communicate in ISL and other people who communicate in Audio or Speech [9]. This gap results in a lack of effective communication in educational institutions, healthcare facilities, and other areas of interaction. Because of the absence of an audio-to-ISL conversion tool, deaf people face challenges when accessing any type of information and getting associated with general society. Expected Solution: The expected solution for this issue includes the development of technology in the form of software or gadgets that can convert speech into ISL. This technology should work on speech recognition, natural language processing (NLP), and computer vision for transcribing and analyzing audio input and providing corresponding ISL signs on a real-time basis [10]. This technology should focus on providing ease of use, accuracy, and the ability to adapt to regional ISL variations, thus enabling smooth interaction and communication among deaf people and those who do not know ISL and communicate through speech.

For instance:

1. Development of an application through which announcements or text on the railway platform Display Unit can be converted into ISL so that deaf people can view and understand the announcement.
2. Development of a mobile application to understand the signs used by deaf people using a Mobile Camera.

PROPOSED SYSTEM ARCHITECTURE

System Overview

The system described has the capability to translate speech (English/Hindi) to ISL and vice versa, making it convenient for deaf and hearing individuals to communicate with each other effortlessly. This system has made accessibility easier in education, government services, health services, and personal communications [9].

There are two main operating modes.

1. *Audio/text-to-ISL translation*: In audio or text-to-ISL translation, audio inputs are captured by a microphone and translated into text format using a speech recognition module. The text is further translated into the ISL grammar format using NLP. ISL grammar is, in the end, translated into visual gestures in the form of animated characters or videos so that deaf individuals can understand it [10].

2. *ISL to audio/text translation*: Images are captured using a camera, which captures the gestural and facial expressions of the user. The processed images from the vision component are analyzed using deep learning algorithms to identify the signs and sequences of those signs. The signs are converted to text or audio for non-sign language users to interpret ISL.

The focus of the system lies in real-time processing, adaptability for regional ISL variations, and high accuracy. The system further includes Web 3.0 techniques for secure and verifiable methods of decentralized storage of gesture datasets and AI models in a manner that does not impose any performance limitations on scalability.

The framework is modular, making it convenient to update each component, such as models, gesture databases, and systems for visualizations, to make it more robust and adaptable to current standards and requirements in ISL.

Module Description

The system comprises the following modules, integrated:

Speech Recognition Module

- It is able to record voice input in English or Hindi.
- It translates speech into text format using an efficient speech-to-text algorithm that can filter and adapt to different accents.
- Ensures accurate transcriptions for proper subsequent gesture creation.

NLP Module

- Decodes transcribed text for semantic interpretation.
- Does things like tokenization, simplification of sentences, and grammatical reconstruction according to the syntax of ISL.
- It retains integrity in representing the gestures.

ISL Gesture Generation Module

- The translated text from the processed text is used to perform ISL gestures with the help of a gesture dictionary and artificial intelligence models.
- Creates animated avatars or video clips to represent ISL. It makes use of various varieties of ISL.
- Gesture recognition-computer vision module: Records hand gestures from camera input.
- It applies deep learning techniques, including convolutional and recurrent neural networks, to identify complex patterns in signs.
- These were translated into signed language messages, which were then analyzed to detect recognizable signs for conversion into text or speech.

Web 3.0 Module

Data Security and Ownership

- Stores ISL datasets, AI models, and user content in a decentralized storage system such as InterPlanetary File System (IPFS).
- This enables the recording of user interaction and contribution of gestures on a blockchain for immutability and transparency purposes.
- Provides mechanisms for Decentralized Identifier (DID) solutions to provide privacy and secure access, where the individual can control their data and contribute to shared gesture libraries.

User Interface Module

- User-friendly interfaces for both the deaf and hearing.
- Provides immediate translation display, rendering of avatars, and feedback.
- Supports execution in browsers, on mobile devices, and on public display systems for announcement or accessibility-related purposes.
- It enables personalization of the avatar display style, gesture speed, and regional variations for the ISL.

METHODOLOGY

The proposed methodology encompasses the procedure involved in implementing a real-time, bidirectional translation platform between spoken language (English or Hindi) or audio, Indian Sign Language, and the translation of ISL into audio or text safely.

Audio

Audio Capture

The audio input is obtained from a microphone, possibly from the user or a public announcement. Other preprocessing techniques that are used to filter signals and prepare them are noise reduction, normalization, and filtering.

Speech Recognition

- More complex models such as Deep Speech or Whisper are used to translate audio into text format.
- Variations in accent, speech rate, and background noise were addressed to facilitate transcription accuracy.

Natural Language Processing

- The transcribed text goes through tokenization, syntax analysis, and semantic parsing cycles.
- The text is processed into a structured form that is amenable to ISL grammar and context understanding.
- Sentence simplification and handling regional variants of ISL vocabulary are also carried out at this level.

Gesture Generation by ISL

- The structured text is then translated into the corresponding ISL gestures through a process facilitated by a gesture dictionary and artificial intelligence models.
- Gestures are represented through animated avatars and videos, making it easier to understand and interpret sign language.
- This module incorporates meaning from context and enables proper communication beyond the translation level of word to sign.

Indian Sign Language

Gesture Capture: Input Device

- The input device used in this project was a camera, which captured hand movements, facial expressions, and other body actions.
- Video preprocessing helps remove background noise and amplify or highlight gestures.

Gesture Recognition:

- For the detection and recognition of gestures, deep learning techniques, such as convolutional neural networks for spatial information and recurrent neural networks or long short-term memory for temporal information, are employed.
- Gesture sequences were analyzed to analyze the sentence, not the sign itself.

Text/Audio Generation

The identified gestures corresponded to organized texts. The synthesized voice or text is used for the deaf group and provides instant comprehension of the ISL discussion for the hearing community.

Data Handling and Security

Dataset Repository

Datasets with gestures, AIs, and user contributions are stored in decentralized storage solutions called IPFS. This ensures highly reliable and evenly distributed data management.

Blockchain Verification

- Blockchain offers an irreversible ledger containing updates and operations, including gesture datasets, AI models, and contributions from users.
- Smart contracts are responsible for the administration of access permissions and the operations of transparency and security.

User Ownership and Privacy

- Decentralized Identity solutions allow the user to have control over their data.
- Personal contributions, such as examples of gestures, can be safely linked only to the user.

INTEGRATION OF WEB 3.0 TECHNOLOGY

The integration of Web 3.0 technology into the proposed ISL translation system ensures that data security, transparency, and user ownership are superior to those of traditional centralized systems. Web 3.0 technologies, specifically blockchain, IPFS, and DID, among other supporting technologies, ensure that datasets, AI models, and user-generated content are trusted, tamper-proof, and owned by the users themselves [6].

Decentralized Data Storage

Gesture datasets, AI models, and system resources are maintained through decentralized storage platforms, such as the Interplanetary File System. Unlike traditional server setups with a single point of failure, decentralized storage scatters data across hundreds of nodes, dramatically reducing the possibility of data tampering. Therefore, gesture libraries, translation models, and user contributions are publicly accessible, verifiable, and permanently stored without dependency on a central authority [6].

Data Integrity and Verification Using Blockchain Technology

Blockchain assists in keeping the updates in the dataset open and immutable. For instance, when there are changes in the models or the history of interactions among users, they are trackable. Every action or change in the models is considered to be an action in the blockchain system, which makes it difficult for changes in the records to occur in the future without the consensus of most people [6].

Access Control Using Smart Contracts

The access rules for gesture data sets, artificial intelligence models, and other resources are programmed in smart contracts. Smart contracts simply enforce access rules in a clear and consistent manner, granting access only to authorized individuals or applications and tracking such events for auditing. Such a system ensures that any private information, such as gestures created by users, cannot be misused [6].

Decentralized Identity for User Ownership

Users are rewarded with decentralized identities when they contribute to predefined gestures or interact with the system. DID allows individuals to retain control and ownership of their data while contributing to shared datasets. This fosters privacy, reduces the chances of misuse of personal contributions, and requires outright user consent for every action performed with the data.

Why Web 3.0 Fits Well?

Security: Data resides in decentralized storage, shielded by blockchain, thus making tampering and breaches far more difficult.

- *Transparency*: Any change or contribution is entered into the blockchain and is thus publicly viewable.
- *User ownership*: Users retain ownership and control over all their data, including gestures, translations, and customized inputs.
- *Scalability*: It can accept more and more datasets and users without overloading a single central server.
- *Reliability*: Distributed storage provides continued availability and fault tolerance, even if some nodes fail.

USE CASE SCENARIO

The proposed AI-based ISL translation system can be applied in multiple real-world scenarios, enabling seamless communication between deaf and hearing users. The following use cases demonstrate the practical applications and benefits of this system.

Educational Establishments

Scenario

The classroom has students who are both deaf and hearing. Lectures, announcements, and discussions are presented in English or Hindi.

Application

Audio lectures are captured and simultaneously translated into ISL by the system [3].

ISL translation appears for deaf students on their devices or classroom displays as animated avatars or video sequences [4, 5].

Questions in ISL from deaf students can be translated into text or spoken formats for teachers and hearing peers. Advantage: This helps deaf students fully participate in learning opportunities.

Public Services and Transportation

Scenario

Train stations, airport terminals, and government agencies are places where critical announcements may be made orally, making it difficult for deaf individuals to interpret them.

Applications in the System

Real-time announcements are recorded and converted into ISL, which is then projected on public screens in the form of avatars or videos. Deaf people can also use kiosks or service counters in ISL, whose gestures can then be converted into text or voice for hearing people [3, 4].

Benefit: This assists in making public spaces more deaf-friendly, thus promoting deaf independence and deaf safety [2].

Healthcare and Hospitals

Situation: Deaf patients have to be able to comprehend symptoms, understand medical instructions, and comprehend medical information.

System Application

The system translates verbal communication by healthcare providers into ISL. Signs of deaf patients are identified, and their meanings are conveyed in text or speech [3].

Benefit

Quality of care and/or patient safety.

Personal Communication

- *Scenario*: Deaf and hearing people meet in homes, working environments, or gatherings.
- *System application*: Conversations, video meetings, and messages are translated from ISL to/from other languages in real-time. Avatars, gestures, and translation speeds can be customized [4].
- *Benefit*: Facilitates effective and natural communication and contributes to overcoming isolation caused by deafness.

Contribution to Community and Research

The gesture information and inputs are stored in a secure, decentralized system so that researchers can increase the ISL gesture set by adding new commands. The real-time translation capabilities of ISL are also flexible for different variants, thereby promoting cultural and linguistic diversity [1, 6].

CONCLUSION

This project demonstrates an adaptive, real-time, two-way system that connects English speech with Hindi and ISL.

It combines artificial intelligence speech recognition capabilities with computer vision, incorporating concepts from Web 3.0 to provide users with safe and genuine interactions within their own realm.

In this proposal, we address some of the major weaknesses in modern sign language translation systems, including vocabulary and regional flexibility, real-time processing challenges, and concerns related to both processing and privacy. By employing deep learning techniques for gesture and language interpretation, we aim to promote high relevance and accuracy in translations. Video gesture information and animated avatars enhance sign language interpretation, while text and speech features enhance accessibility.

This is where artificial intelligence intersects with decentralized technology to make data more secure, understandable, and owned by all. By using decentralized storage, blockchain, and decentralized identity, we can enhance gesture libraries to make models more intelligent. When AI meets Web 3.0, you have a technology you can trust. Looking ahead to ISL, a couple of scenarios are beginning to emerge—extending the existing collection of ISL gestures to include regional and contextual nuances.

As a bonus, mobile app solutions and public display solutions that anyone can use can be developed. All of this is a critical part of a deaf and hard-of-hearing group's shortened gap by integrating artificial intelligence solutions with a decentralized systems framework.

REFERENCES

1. Hussain A, Saikia N, Dev C. Advancements in Indian sign language recognition systems: enhancing communication and accessibility for the deaf and hearing impaired. *Asian J Electr Sci*. 2023;12(2):37-49. doi:10.51983/ajes-2023.12.2.4132.
2. Sridhar A, Poddar R, Jain M, Kumar P. Challenges faced by the employed Indian DHH community. In: *Proceedings of the 19th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2023)*; 2023 Aug 28-Sep 1; York, UK. Cham: Springer; 2023. p.201-23. doi:10.1007/978-3-031-42280-5_13.
3. Jurafsky D, Martin JH. *Speech and language processing*. New Delhi: Pearson Education India; 2000.

4. Bantupalli K, Xie Y. American sign language recognition using deep learning and computer vision. In: Proceedings of the 2018 IEEE International Conference on Big Data (Big Data); 2018 Dec 10-13; Seattle, WA. p.4896-9. doi:10.1109/BigData.2018.8622141.
5. Starner T, Weaver J, Pentland A. Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans Pattern Anal Mach Intell. 1998;20(12):1371-5. doi:10.1109/34.735811.
6. Wood G. Ethereum: a secure decentralised generalised transaction ledger [Internet]. Ethereum Project Yellow Paper; 2014.
7. Tikhomirov S. Ethereum: state of knowledge and research perspectives. In: Imine A, Fernandez JM, Marion JY, Garcia-Alfaro J, Logrippo L, editors. Foundations and Practice of Security. Cham: Springer International Publishing; 2017. p.206-21. doi:10.1007/978-3-319-75650-9_14.
8. Antonopoulos AM, Wood G. Mastering Ethereum: building smart contracts and DApps. Sebastopol (CA): O'Reilly Media; 2018.
9. Das S, Biswas SK, Chakraborty M, Purkayastha B. Intelligent Indian sign language recognition systems: a critical review. In: ICT Systems and Sustainability. Singapore: Springer; 2022. p.703-13. doi:10.1007/978-981-16-5987-4_71.
10. Raj R, Bhat A. Insights into Indian sign language recognition: a comprehensive review. In: World Congress on Smart Computing. Singapore: Springer Nature Singapore; 2025. p.251-62. doi:10.1007/978-981-97-9006-7_21.