International Journal of Engineering Research & Science (IJOER) ISSN: [2395-6992] [Vol-12, Issue-3, March 2026]

RAG and LSTM Based Communication Aid for the Deaf and Silent Community

Archana Ingle[1*]; Bishal Dubey[2]; Sahil Prajapati[3]; Amit Yadav[4]

Department of EXTC, VIVA Institute of Technology / University of Mumbai, India
*Corresponding Author

Abstract— *Communication barriers between the hearing population and individuals with hearing and speech impairments remain a major societal challenge in education, healthcare, and social inclusion. This paper presents a bidirectional assistive communication system that translates spoken/written language into Indian Sign Language (ISL) and vice versa. The Speech-to-Sign pipeline employs Retrieval-Augmented Generation (RAG) with FAISS-based semantic search to map sentences into contextually accurate ISL video clips. The Sign-to-Speech pipeline integrates MediaPipe-based keypoint extraction with sequence models such as LSTM and Transformers to recognize dynamic gestures in real time. FAISS-based semantic search combined with Natural Language Processing (NLP) enables accurate retrieval of relevant ISL sign videos even for paraphrased inputs. In the reverse translation pipeline, MediaPipe hand landmark detection and LSTM-based sequence modeling allow real-time gesture recognition through a webcam interface*.

Keywords— Sign Language Recognition (SLR), Indian Sign Language (ISL), Speech-to-Sign Translation, Sign-to-Speech Translation, Assistive Technology, Deep Learning, MediaPipe, LSTM, Transformer, Retrieval-Augmented Generation (RAG), Human–Computer Interaction (HCI).

I. INTRODUCTION

Communication is central to human interaction, allowing people to share thoughts and emotions. For individuals with hearing and speech impairments, sign language is the primary mode of expression. In India, Indian Sign Language (ISL) is widely used, but most hearing individuals do not understand it. This lack of mutual understanding creates barriers in education, healthcare, employment, and daily social life, often leading to exclusion and reduced opportunities for the deaf and silent community. Bridging this gap is therefore not only a technical task but an important social responsibility [1].

Many existing assistive systems attempt to solve this problem, but they usually depend on costly and bulky hardware such as sensor gloves or motion tracking devices. Several research efforts also concentrate mainly on American Sign Language (ASL) or British Sign Language (BSL), leaving ISL comparatively underexplored. As a result, ISL resources are limited, and present systems often struggle with context handling, real-time performance, and mobile deployment. Word-level translation misses grammatical structure, and high computational demand restricts use on common devices.

To overcome these issues, this paper presents an AI-powered bidirectional communication system that enables interaction between ISL users and non-signers. The first module translates speech or text into ISL videos using a Retrieval-Augmented Generation approach, where semantic embeddings and similarity search help select appropriate gesture clips. The second module converts ISL gestures into speech by using MediaPipe keypoints with LSTM and Transformer-based sequence models for real-time recognition. Together, these modules support communication in both directions.

The proposed system emphasizes cultural accuracy, lightweight architecture, and real-time response. Instead of synthetic avatars, it uses curated ISL video datasets, resulting in natural and meaningful gestures. The modular design also allows expansion to other sign languages and larger vocabularies in the future [2], [3].

The main objectives are to develop an affordable, portable ISL translation tool, provide low-latency real-time interaction, improve contextual accuracy, and promote social inclusion. The key contributions include a novel dual-direction framework, curated ISL dataset creation, robust gesture-recognition pipeline, and comprehensive performance evaluation.

II. LITERATURE REVIEW

The field of assistive communication technologies for Deaf and Hard-of-Hearing individuals has witnessed remarkable growth in recent years, driven by the urgent need to bridge the communication gap between sign language users and the hearing population. In India, Indian Sign Language (ISL) serves as the primary medium of communication for thousands of individuals, yet the majority of the general population remains unfamiliar with it. This lack of fluency creates barriers across multiple domains such as education, healthcare, employment, and social participation [1].

Early attempts to address this challenge relied heavily on hardware-based solutions. Devices such as glove sensors, accelerometers, and motion trackers were used to capture hand movements with high precision. These systems demonstrated technical feasibility but were limited by their cost, bulkiness, and lack of portability [2]. As technology advanced, researchers began to explore camera-based gesture recognition systems. Computer vision techniques, particularly Convolutional Neural Networks (CNNs), were employed to classify static gestures, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks were introduced to capture temporal dependencies in dynamic gestures. Retrieval-based methods also emerged, with semantic search and Retrieval-Augmented Generation (RAG) being adapted to select contextually appropriate video sequences [3].

Deshmukh et al. [4] introduced a real-time communication platform that converted spoken English into ISL gestures using speech recognition and natural language processing (NLP). The system was able to process speech over short time windows and generate synchronized sign videos. Despite its strengths, the system faced challenges due to limited datasets and difficulties in handling background noise and varied accents.

Parivazhagan et al. [5] presented a web-based platform that converted spoken English into ISL using speech recognition APIs, NLP preprocessing, and avatar-based animations. Rule-based grammar transfer was applied to rearrange English syntax into ISL grammar. While the integration of advanced NLP techniques improved grammatical accuracy, limitations such as low-resolution animations and restricted vocabulary reduced its effectiveness.

Rouf and Jadon [6] integrated generative artificial intelligence for text-to-speech and sign language translation. These systems combined automatic speech recognition, NLP, and three-dimensional avatar animation to create dynamic sign language outputs. Models such as Whisper and Google Cloud ASR achieved high accuracy in controlled environments.

Shofia Priyadharshini et al. [7] developed a comprehensive application for sign language alphabet and word recognition, text-to-action conversion, multi-language support, and voice output. While the system promoted inclusivity, weaknesses included limited gesture datasets, hardware dependency, and performance constraints on low-end devices.

Venkatesh et al. [8] proposed a quantum convolutional neural network (QCNN)-based real-time sign language detection system deployed on Raspberry Pi. Although QCNN integration improved recognition precision, challenges such as hardware constraints and environmental sensitivity limited scalability.

Sathishkumar et al. [9] developed a sign language detection system using action recognition techniques, demonstrating the feasibility of vision-based approaches for continuous sign recognition.

Gusain et al. [10] implemented real-time sign language recognition using MediaPipe and random forests, showing the effectiveness of MediaPipe for hand landmark extraction.

The existing literature highlights that most systems lack bidirectional communication and are constrained by limited datasets and language coverage. These limitations indicate the need for an intelligent, scalable, and real-time bidirectional framework, motivating the development of the proposed system.

III. MATERIAL AND METHODS

The proposed system is designed as a bidirectional assistive communication framework that enables interaction between hearing individuals and users of Indian Sign Language (ISL). The overall methodology consists of two complementary pipelines: (1) Speech/Text to ISL Sign Video, and (2) Sign Gesture to Text/Speech Output.

3.1 Speech/Text to Sign Language Module

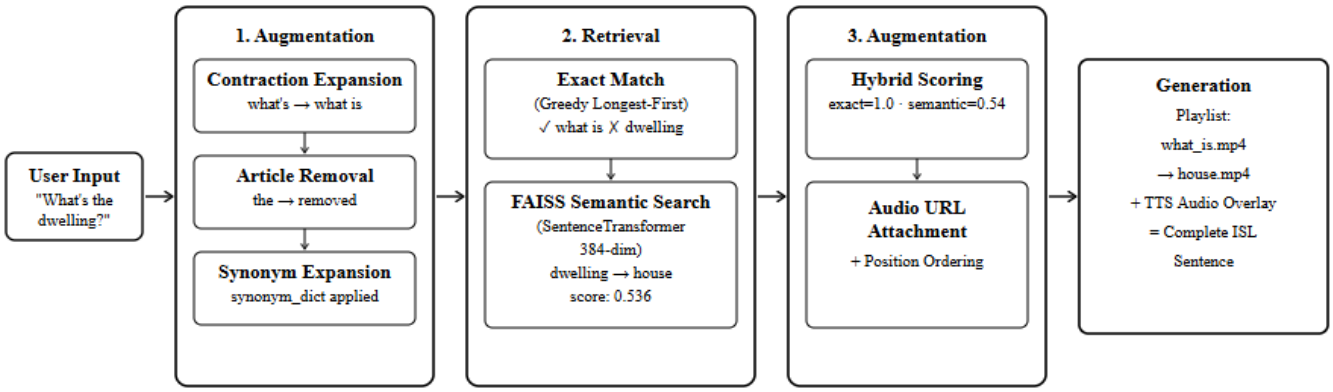

Figure 1: Text/Speech to Hand Sign Language Flowchart

As shown in Figure 1, the input is first obtained as raw text or through speech-to-text transcription. The text passes through a preprocessing layer that removes punctuation, converts words to lowercase, and performs synonym normalization, ensuring that different phrasings expressing the same meaning are treated uniformly.

The processed text is represented as dense semantic embeddings using a transformer-based sentence encoder. These embeddings capture contextual meaning rather than relying only on keywords. A vector similarity search engine built on **FAISS** is used to compare the query embedding with stored embeddings of ISL phrases. A multi-stage matching strategy is applied, combining exact phrase lookup, fuzzy similarity matching, semantic similarity retrieval, and synonym-based fallback resolution. Once the closest match is determined, the corresponding human-recorded ISL video is retrieved from the database and presented to the user.

3.2 Sign Language to Speech/Text Module

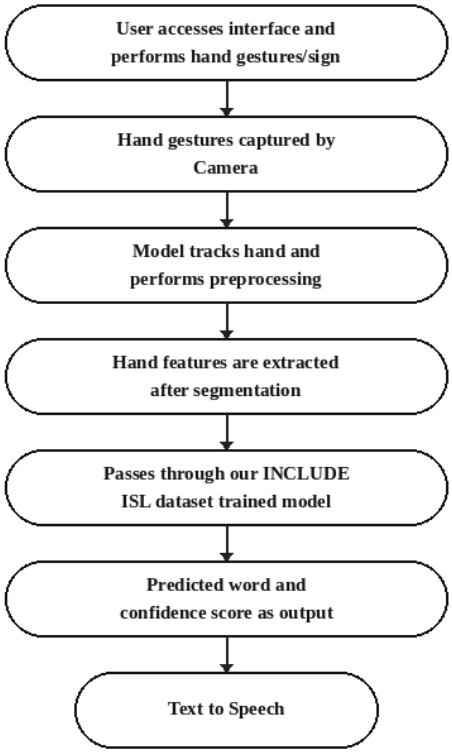

Figure 2: Sign Language to Text/Speech Flowchart

For the reverse translation direction, the system acquires video input through a front-facing camera operating at standard frame rates. Using **MediaPipe**, detailed landmarks for hands, face, and body are extracted as shown in Figure 2. These keypoints are normalized and converted into temporal descriptors such as joint angles, motion trajectories, and velocities.

The time-varying feature sequence is then processed by deep sequence models such as **Long Short-Term Memory (LSTM)** networks or **Transformer-based encoders**, which learn relationships across consecutive frames. The network outputs either gloss sequences or textual representation of the signed phrase. A language-model-based post-processing step improves grammatical structure and readability. The final recognized text may be optionally converted into audio using text-to-speech synthesis.

The system is implemented as a modular framework, enabling easy extension of vocabulary and addition of new sign videos without modification of the core algorithm.

IV. RESULTS AND DISCUSSION

The proposed bidirectional communication system was experimentally evaluated to assess its performance in real-life scenarios. The system was tested for two major functionalities: (1) conversion of speech/text into Indian Sign Language (ISL) videos, and (2) recognition of sign language gestures and conversion into text and speech output. A dataset consisting of commonly used daily-life gestures and conversational sentences was used for performance evaluation, including greetings, self-introduction statements, basic question words, and frequently used actions.

4.1 Speech/Text to Sign Language Module

Figure 3: Text/Speech to Hand Sign Language

The system successfully converted input sentences into corresponding ISL videos by using semantic similarity-based retrieval. Even when the input was paraphrased or grammatically varied, the retrieval mechanism was able to identify the closest matching sign video. The results show that the system is capable of handling synonyms and minor sentence variations without significant loss of meaning.

Figure 3 illustrates the system converting the text "Whatsup" into its corresponding Indian Sign Language (ISL) hand gesture. The input text is processed through an NLP module, which performs tokenization and semantic analysis. The RAG-based retrieval pipeline then selects the most relevant gesture from the FAISS database, demonstrating the system's ability to handle common single-word phrases in near real-time.

4.2 Sign Language to Text Module

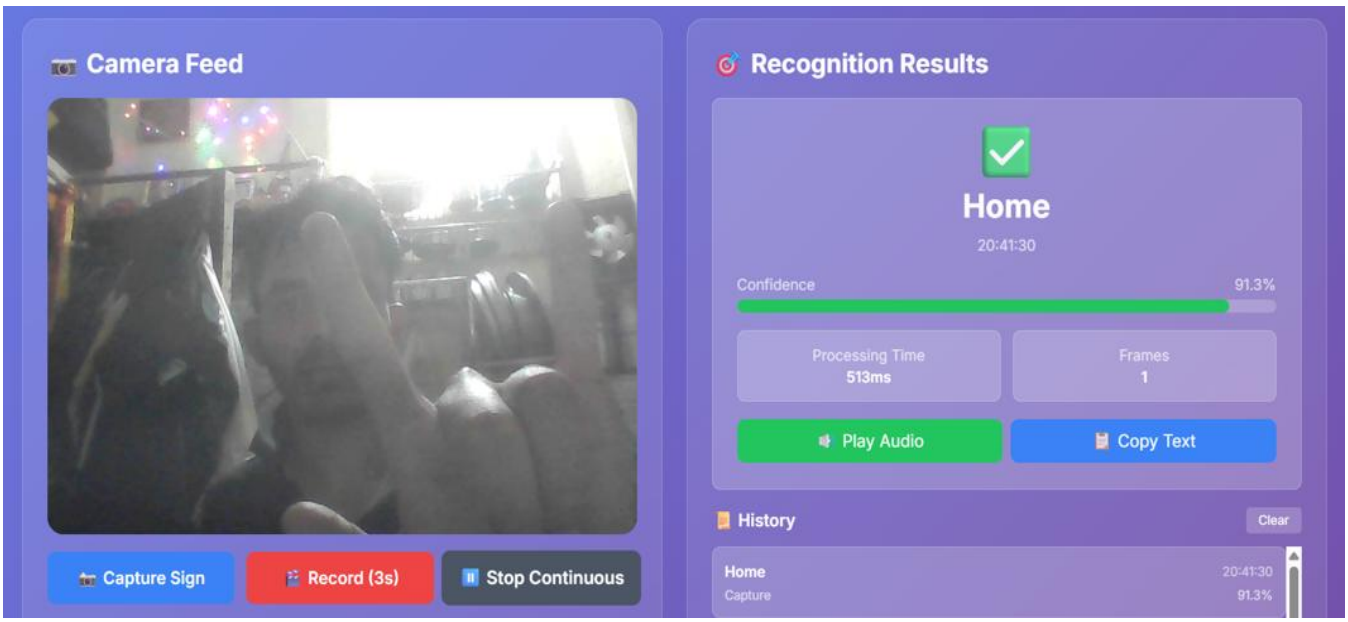

Figure 4: Hand Sign Language to Text/Speech

Figure 4 shows the web interface of the proposed Sign Language Recognition System. The interface is divided into two main sections: Camera Feed and Recognition Results. On the left side, the Camera Feed panel displays the live video stream captured from the user's webcam. The panel includes control buttons such as "Capture Sign," "Record (3s)," and "Stop Continuous."

On the right side, the Recognition Results panel displays the output generated after the gesture is processed. In the example shown, the system successfully recognizes the gesture as **"Home"** with a confidence level of **91.3%**, demonstrating high accuracy. Additional information such as processing time (513ms) and the number of frames used for detection is also displayed.

Below the result, the interface provides interactive options including "Play Audio" (converts recognized text to speech) and "Copy Text" (copies the recognized output). The History section records previously recognized gestures along with timestamps and confidence values.

TABLE 1
PERFORMANCE SUMMARY

Module	Task	Accuracy	Latency
Speech-to-Sign	Text/audio to ISL video	High (semantic retrieval)	Near real-time
Sign-to-Speech	Gesture recognition	91.3% (sample)	~513ms

V. CONCLUSION

This work presents a bidirectional communication system designed to bridge the gap between individuals who use Indian Sign Language (ISL) and those who communicate through spoken language. The system integrates two major functionalities: conversion of speech or text into ISL sign videos, and recognition of sign gestures followed by conversion into text and speech output.

Key Contributions:

1. Novel dual-direction framework combining RAG-based retrieval and LSTM-based gesture recognition
2. Semantic similarity-based matching for paraphrased input handling
3. MediaPipe-based keypoint extraction for lightweight, real-time gesture recognition

4. Curated ISL video dataset for authentic sign representation

Limitations:

- Performance affected by rapid gesture movement, partial hand occlusion, and low-light conditions
- Vocabulary limited to commonly used daily-life expressions
- Continuous sign recognition not yet fully implemented

Future Work:

- Continuous sign language recognition for natural conversation flow
- Expanded ISL vocabulary including complex sentences and non-manual cues
- Mobile and edge-device deployment for improved accessibility
- Multilingual support for other Indian regional languages
- Integration of advanced deep learning architectures for enhanced contextual understanding

ACKNOWLEDGMENT

Special thanks to the research team at the Indian Institute of Technology Madras (IIT Madras) for developing the INCLUDE dataset for Indian Sign Language, which served as an important resource for training and evaluating the gesture recognition model used in this project. The authors also acknowledge the creators and contributors on YouTube whose educational videos and sign language demonstrations were used as references and for the retrieval component of this project.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication and development of this project. The work presented in this report was carried out solely for academic purposes as part of the project requirements. No financial or commercial relationships existed that could have influenced the outcomes or interpretations of the research.

REFERENCES

[1] A. Deshmukh, A. Machindar, S. Lale, and P. Kasambe, "Enhancing communication for the hearing impaired: A real-time speech-to-sign language converter," in Proc. 27th Int. Symp. Wireless Personal Multimedia Communications (WPMC), vol. 27, IEEE, 2024, pp. 1–6

[2] A. Parivazhagan, K. R. Penugonda, N. Pasham, S. Neela Krishna, and T. Mahendra Reddy, "AI-powered real-time speech-to-sign translation," in Proc. 5th Int. Conf. Pervasive Computing and Social Networking (ICPCSN), vol. 5, IEEE, 2025, pp. 1389–1394

[3] K. Shirisha, E. M. Deeksith, M. S. S. Sri Madhava, and R. Karthikeyan, "SIGNBRIDGE: Audio to sign language translator," in Proc. IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS), vol. 1, IEEE, 2025, pp. 1–5

[4] M. Rouf and J. S. Jadon, "Generative AI for text-to-speech and sign language translation," in Proc. IEEE Int. Conf. on Data Technology (ICDT), vol. 1, IEEE, 2025, pp. 1124–1129.

[5] D. Shofia Priyadharshini, R. Anandraj, K. R. Ganesh Prasath, and A. Ganesh, "A comprehensive application for sign language alphabet and word recognition, text-to-action conversion for learners, multi-language support and integrated voice output functionality," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2024, pp. 101–108.

[6] B. Venkatesh, D. Nagajyothi, P. Dheeraj Kumar, and K. Ravi Teja, "Real-time sign language translation system using QCNN for enhanced accessibility," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2024, pp. 210–216.

[7] C. Bhat, R. Rajeshirke, S. Chude, V. Mhaiskar, and V. Agarwal, "Two-way communication: An integrated system for American sign language recognition and speech-to-text translation," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2023, pp. 55–61

[8] P. Sathishkumar, V. H. Iyer, U. M. Prakash, and A. Vijay, "Sign language detection using action recognition," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2022, pp. 89–94.

[9] P. Gusain, A. Verma, and G. Sharma, "Real-time sign language recognition and translation using MediaPipe and random forests for inclusive communication," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2024, pp. 301–307

[10] P. Goriparthi, A. Pandiaraj, and P. Nancy, "Bridging the gap: Real-time Telugu and Tamil sign language recognition using AI and computer vision," in Proc. IEEE Int. Conf., vol. 1, IEEE, 2025, pp. 415–421.