



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.935

Volume 15, Issue 10, October 2026

SanvaadISL: A Bidirectional Indian Sign Language Communication System using Deep Learning

Prof.N.B.Gade, Manasi Walhekar, Priyanka Andhale, Vaishnavi Nagpure, Chaitanya Rohokale

HOD, Department of Computer, SCSMCOE, Maharashtra, India

Department of Computer, SCSMCOE, Maharashtra, India

ABSTRACT: Communication can be difficult for deaf and hard-of-hearing people when the people they need to interact with do not understand Indian Sign Language (ISL). The challenge is especially noticeable in places such as banks, hospitals, post offices, police stations, and government service centres, where clear communication is essential. This paper presents SanvaadISL, a proposed bidirectional ISL communication application that combines computer vision, landmark extraction, gesture recognition, text-to-speech, speech-to-text, and sign-video retrieval. In the first direction, a camera captures a user's signs, MediaPipe extracts relevant landmarks, and a trained recognition model identifies signs within its supported vocabulary. The result is displayed as text and may also be spoken aloud. In the reverse direction, English or Hindi text—or text produced from speech—is matched with entries in a curated vocabulary, and the corresponding pre-recorded ISL sign video is displayed. A key focus is domain-specific vocabulary for everyday service interactions, including common questions, requests, and responses. The system will be evaluated using accuracy, precision, recall, F1-score, response time, and task-based tests. Its effectiveness will depend on implementation and experimental results; the paper therefore treats performance as an evaluation objective rather than an established outcome.

KEYWORDS: Indian Sign Language, bidirectional communication, MediaPipe, deep learning, gesture recognition, sign-video retrieval, accessibility, domain-specific vocabulary.

I. INTRODUCTION

Communication plays an important role in accessing healthcare, education, banking, and public services. Deaf and hard-of-hearing people who use Indian Sign Language may face difficulties when staff members or other people do not understand ISL. These difficulties can make routine tasks—such as explaining a medical concern, asking for a bank statement, checking the status of an Aadhaar update, or requesting help at a police station—more complicated. The limited availability of interpreters in everyday settings makes accessible communication tools worth exploring.

Advances in computer vision and deep learning have made camera-based sign recognition increasingly practical. Such systems can process video frames, identify useful visual features, and classify signs using trained models. MediaPipe can extract hand, body, and facial landmarks, while sequence models can learn how those landmarks change during a gesture. Prior work has explored CNNs, recurrent neural networks, gated recurrent units, and hybrid approaches for sign recognition and for generating text or speech [1]–[4]. However, performance can vary with the dataset, vocabulary size, signing style, lighting, background, and complexity of the gestures.

Recognizing a sign is only one part of a useful communication system. People also need a way to respond to a signer, ask follow-up questions, and exchange information about a particular service. A general-purpose vocabulary may not contain expressions commonly used in banks, hospitals, post offices, police stations, Aadhaar Seva Kendras, income tax offices, and other government service centres. SanvaadISL is proposed to address this practical need by combining sign-to-text/speech recognition with a reverse text-or-speech-to-sign-video pathway and a collection of service-specific expressions.

The main contribution of this work is the proposed integration of two communication directions and domain-specific vocabulary in one interface. The reverse pathway retrieves pre-recorded videos for supported words and phrases; it

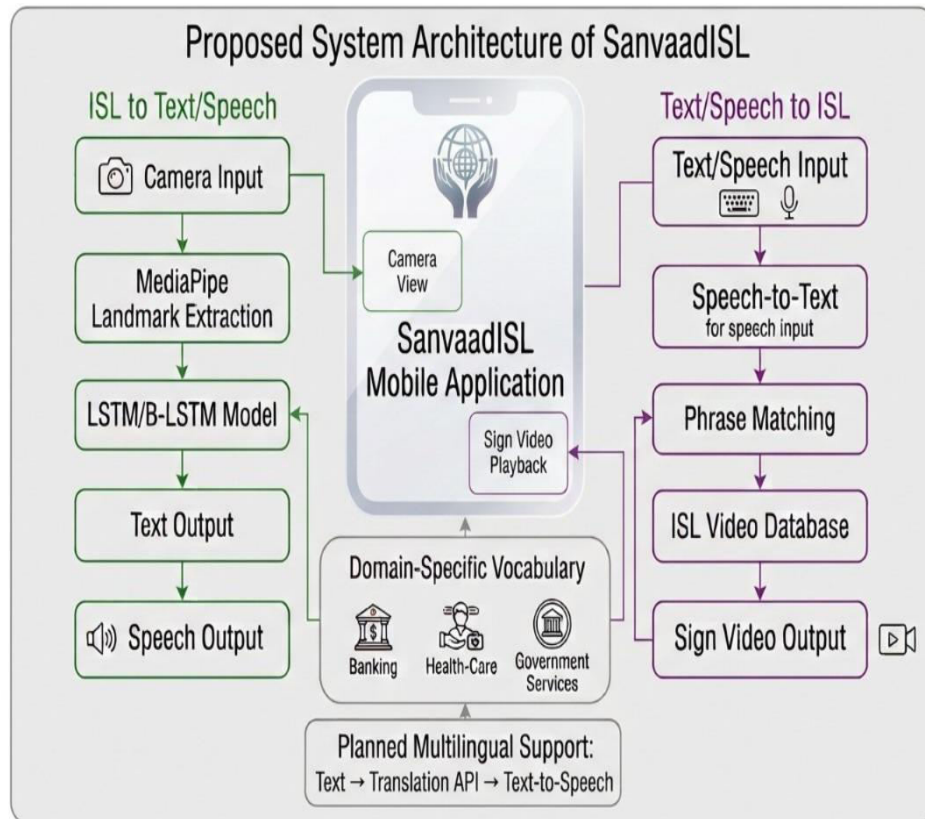
should not be interpreted as unrestricted translation of arbitrary English or Hindi sentences into grammatically complete ISL. The vocabulary and sign-video mappings should be reviewed by competent ISL users or interpreters before deployment.

II. LITERATURE SURVEY

Existing studies provide useful building blocks for sign-language communication, including real-time gesture recognition, sequential modelling, and text or speech output. The studies summarized below differ in language, dataset, and evaluation setup. Their reported results should therefore not be compared directly unless they were obtained under equivalent experimental conditions.

Author and year	Method / architecture	Dataset / language	Accuracy results	Main contribution
Dipon Talukder and Fatima Jahara (2020) [1]	Real-time sign detection with sentence generation and speech output.	Bangla Sign Language (BdSL).	A numerical accuracy value was not verified from the accessible source.	Extends recognition beyond individual sign labels by exploring sentence and speech generation.
Barathi Subramanian et al. (2022) [2]	MediaPipe Holistic with a modified GRU architecture called MOPGRU.	A collected set of 13 ISL gestures; also evaluated on WLASL100 (ASL) and LSA64 (Argentinian Sign Language).	63.18% on WLASL100 and 99.92% on LSA64, as reported in the supplied material. These results come from different datasets.	Modifies GRU gating and activation functions for sequential gesture recognition.
B. Surya, N. V. Suresh Krishna et al. (2023) [3]	CNN-based recognition with MediaPipe and OpenCV for detection, tracking, segmentation, and feature extraction.	Indian Sign Language; the accessible abstract does not confirm a benchmark dataset or vocabulary size.	A numerical accuracy value was not verified from the accessible abstract.	Combines webcam-based ISL recognition with text-to-speech and other accessibility features.
Malisetty Revathi, Karthik Kuncha, Medisetty Shanmukha Nanda et al. (2025) [4]	CNN-LSTM-based real-time sign recognition with text and speech output.	The exact training dataset and vocabulary should be checked against the full paper.	A numerical accuracy value for the paper's model was not verified from the supplied metadata.	Connects real-time sign recognition with text and speech output.
Proposed SanvaadISL	MediaPipe landmark extraction and a trained gesture-recognition model, integrated with text/speech output and sign-video retrieval.	ISL, with planned vocabulary for banks, hospitals, post offices, police stations, Aadhaar Seva Kendras, income tax offices, government services.	To be measured experimentally using accuracy, precision, recall, F1-score, latency, and task-based evaluation.	Combines two-way communication with planned domain-specific vocabulary and service-oriented workflows.

III. ARCHITECTURE



The proposed system, SANVAADISL, enables two-way communication between Indian Sign Language (ISL) users and non-signers through two pipelines. Pipeline A captures sign gestures using a camera, preprocesses video frames with OpenCV, extracts hand landmarks using MediaPipe, and recognizes gestures using a trained machine learning or deep learning model to generate text and speech. Pipeline B accepts English/Hindi text or speech, processes the input, maps words to corresponding ISL signs, and retrieves sign videos from a structured database. The system integrates gesture recognition, speech processing, and sign-video retrieval to support accessible and inclusive communication.

IV. PROPOSED SYSTEM

SanvaadISL is proposed as a mobile-oriented application that supports communication between ISL users and people who do not know sign language. It has two main pathways. The first recognizes supported ISL gestures and presents the result as text and optional speech. The second accepts English or Hindi text, or speech converted into text, and retrieves a matching pre-recorded ISL sign video from a structured library. The initial application is intended to support a defined vocabulary rather than unrestricted, sentence-level translation.

1. ISL-to-Text and Speech

The camera captures the user's signing. Video frames are processed, and MediaPipe is used to extract relevant hand landmarks and, where required, body or facial landmarks. The resulting features are passed to a trained recognition model, which predicts one of the supported sign classes. The predicted label is displayed as text and can be sent to a text-to-speech component. Recognition is limited to the signs represented in the training data and supported by the model.

2. Text or Speech-to-ISL Video

In the reverse pathway, a user enters English or Hindi text, or speaks into a microphone. If speech is used, a speech-to-text component converts it into text. The application then matches the text against a predefined set of words and phrases and displays the corresponding sign video when a match is available. If the input is not in the supported vocabulary, the interface should clearly indicate that no matching sign entry was found rather than presenting an incorrect result.

3. Domain-Specific Vocabulary

The vocabulary will be organized into service categories such as banking, healthcare, postal services, police services, Aadhaar services, income tax services, and other government offices. Each category may contain commonly used questions, requests, instructions, and responses. Examples include asking for an appointment, requesting a document, describing a symptom, checking an application status, or asking where to submit a form. The vocabulary and its sign-video mappings should be validated with ISL users or qualified interpreters to help ensure that the content is accurate and understandable.

V. SYSTEM WORKFLOW

1. Data Preparation

The dataset will be prepared from suitable publicly available ISL resources and, where necessary, additional recorded samples. Each sample will be labelled with the corresponding sign or phrase. MediaPipe landmarks can then be extracted and arranged into sequences for model training. A possible split is 70% training, 15% validation, and 15% testing. The split should be designed to avoid data leakage; where possible, samples from the same signer or recording session should not appear across training and test sets.

2. Model Design and Training

Landmark sequences will be used to train a sequence model such as LSTM or BiLSTM to classify the supported gestures. The model learns the relationship between the input sequences and their labels. Validation data will support model selection and tuning, while the test set will be kept separate for final evaluation. The final architecture should be selected based on measured performance, not assumed accuracy.

3. Application Development

The application will connect the recognition model with the user interface, text output, optional speech output, text or speech input, and sign-video retrieval. For the forward pathway, camera frames are processed and converted into recognized text. For the reverse pathway, typed or transcribed input is matched against the supported vocabulary and the relevant sign video is retrieved. Domain-specific entries will be grouped into service categories to make common expressions easier to find.

4. Interface Design

The interface will provide clear controls for camera input, text entry, speech input, recognized-text display, speech playback, and sign-video playback. Service categories will help users locate relevant phrases. Multilingual translation through the Google Cloud Translation API and language-specific text-to-speech support may be added as extensions; these features will require implementation, configuration, and testing before they can be described as available.

5. Testing and Evaluation

The model will be tested on samples that were not used during training. Recognition will be evaluated using accuracy, precision, recall, and F1-score. Response latency will be measured to understand how quickly the system produces an output. Task-based tests will assess whether users can complete predefined service-related exchanges using the supported vocabulary. Further tests should consider different signers, signing speeds, lighting conditions, backgrounds, and camera positions. The results will be used to identify limitations and guide future improvements.

VI. TECHNOLOGIES

1. MediaPipe Holistic: Extracts relevant hand, body, and facial landmarks from camera input. These landmarks provide numerical features that can be used by the recognition model.
2. LSTM / BiLSTM: Processes sequences of extracted landmarks to recognize supported ISL gestures. LSTM models temporal dependencies, while BiLSTM processes a sequence in both directions. The final model should be chosen after experimental comparison.
3. Speech-to-Text and Text-to-Speech: Speech-to-text converts spoken input into text for matching, while text-to-speech reads recognized text aloud for hearing users.
4. Text Matching and Sign-Video Retrieval: Matches typed or transcribed input with a predefined vocabulary and retrieves the corresponding pre-recorded ISL video.
5. Multilingual Translation: A planned translation service may convert text into additional supported languages. A compatible speech service may then produce audio in the selected language. These capabilities should be reported as implemented only after they have been integrated and tested.

These technologies support all the functions carried out in the system such as gesture Recognition ,Text to Speech and Speech to Text conversion ,and also predefined sign-video retrieval in the Proposed application.

VII. EVALUTION PLAN

The system's performance will be assessed using classification metrics and practical communication tests. Accuracy will measure the proportion of correctly classified test examples. Precision, recall, and F1-score will provide additional information about class-level performance, especially where the dataset is imbalanced. Response latency will measure the time taken to produce recognized text or retrieve a sign video. Task-based evaluation will record whether users can successfully complete selected service-related interactions. The paper should report the dataset size, number of sign classes, test protocol, hardware, and measured results when experiments are complete. No numerical performance claim should be included until it is supported by the actual evaluation.

VIII. RESEARCH GAP

The reviewed work covers important elements of sign-language recognition and accessible output, but the supplied summaries do not establish that all of these features are combined in a single system focused on everyday public-service interactions. SanvaadISL addresses this design opportunity by bringing together ISL gesture recognition, text and speech output, a reverse sign-video retrieval pathway, and a vocabulary organized around specific service settings. Its contribution is the proposed integration and practical focus. Whether this integration improves communication or task completion must be demonstrated through user testing and experimental evaluation.

IX. CONCLUSION

SanvaadISL is proposed as a bidirectional ISL communication application intended to make everyday interactions between ISL users and non-signers easier. The forward pathway uses camera input, landmark extraction, and a trained recognition model to convert supported signs into text and optional speech. The reverse pathway matches English or Hindi text—or text obtained from speech—with a predefined vocabulary and displays the corresponding ISL sign video. Domain-specific categories for banking, healthcare, postal services, police services, and government offices are intended to make the system more useful in practical situations. The system's actual effectiveness will depend on the quality of the dataset, the accuracy of the sign mappings, usability, and the results of formal testing. Future work can expand the vocabulary, investigate continuous sign-sequence recognition using approaches such as Transformer-CTC, include facial expressions and other non-manual markers, and optimize the application for efficient operation on mobile devices

REFERENCES

1. H. Khubchandani & T. Karthick, "Sign language recognition," The 9th International Conference on Information Technology Trends (ITT), Dubai, United Arab Emirates, 2023, 199 - 203. doi: 10.1109/ITT59889.2023.10184252
2. D. Talukder & F. Jahara, "Real-time bangla sign language detection with sentence and speech generation," 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020
3. M. Revathi et al., "Real Time Sign Language Recognition to Text and Speech using CNN based Deep Learning," 2025 4th International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), 2025, pp. 1737–1742, doi: 10.1109/ICIMIA67127.2025.11200705
4. M. J. Hossein and M. S. Ejaz, "Recognition of Bengali Sign Language using Novel Deep Convolutional Neural Network," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), Dec. 2020, doi: 10.1109/STI50764.2020.9350418.
5. P. Unkule et al., "CNN based Approach for Sign Recognition in the Indian Sign language," 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), pp. 92–97, IEEE, 2022
6. L. Pigou, S. Dieleman, P.-J. Kindermans, and B. Schrauwen, "Sign language recognition using convolutional neural networks," European Conference on Computer Vision (ECCV), pp. 572–578, Springer, 2014.
7. K. Bantupalli and Y. Xie, "American sign language recognition using deep learning and computer vision," 2018 IEEE International Conference on Big Data (Big Data), pp. 4896–4899, IEEE, 2018.
8. "Toward Real-Time Recognition of Continuous Indian Sign Language: A Multi-Modal Approach Using RGB and Pose," IEEE Access, vol. 13, pp. 60270–60283, 2025, doi: 10.1109/ACCESS.2025.3554618.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details