ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

A Deep Learning Approach for Real-time Continuous Indian Sign Language Recognition and Translation

Suraj V. Dhole[1], Nandini Raut[2], Shrushti Shirbhate[3], Radhika Watane[4], Shrushti Wakode[5]

[1]*Assistant Professor, G. H. Raisoni University, Amravati, India*

[2,3,4,5]*Undergraduate Student, G. H. Raisoni University, Amravati, India*

Abstract: *Indian Sign Language (ISL) serves as a vital means of interaction for people with hearing and speech impairments, yet it remains unfamiliar to most non-signers. This gap often results in communication challenges across academic, professional, and social environments. To help address these limitations, this work introduces a real-time ISL translation system that applies computer vision and deep learning techniques to interpret hand gestures and convert them into readable text or synthesized speech. The system captures user gestures through a webcam, extracts hand landmarks using MediaPipe, and classifies them with a Convolutional Neural Network (CNN). A curated dataset of frequently used ISL gestures and alphabets is used to train the model, facilitating accurate recognition and reliable performance in real-time applications.*

Keywords: Indian Sign Language, Deep Learning, Gesture Recognition, Computer Vision, Assistive Technology, Real-Time Processing.

I. INTRODUCTION

Human communication is essential for expressing ideas, emotions, and information. For individuals with hearing or speech impairments, however, interacting with the broader community can be challenging because their primary communication medium—Indian Sign Language (ISL) is not widely understood by others. This lack of familiarity can impede their participation in routine activities such as education, healthcare, and workplaces.

Advances in artificial intelligence and image analysis now make it possible to design systems that can recognize hand gestures and translate them automatically. This project aims to create a real-time ISL translation system capable of converting gestures into text or speech, helping to bridge communication gaps without relying on a human interpreter.

1.1 Motivation

India has a large population of deaf and mute individuals, many of whom depend heavily on ISL for daily communication. However, the availability of trained sign language interpreters remains very limited, restricting equal access to essential services and opportunities. By utilizing recent developments in deep learning and computer vision, this project seeks to reduce communication obstacles and improve accessibility. A system that can interpret ISL gestures autonomously can promote independence, support inclusive communication, and enhance the overall quality of life for the hearing-impaired community.

 DOI: 10.5281/zenodo.17739533

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

1.2 Problem Definition

The core issue addressed in this work is the scarcity of ISL-aware individuals among the general population, which makes communication for the hearing-impaired community difficult and often dependent on intermediaries. Although interpreters exist, they are insufficient in number. This project proposes an automated system that recognizes ISL hand gestures and converts them into text and speech. The goal is to create a real-time translation tool that can assist users in communicating more seamlessly with individuals who do not understand sign language.

1.3 Objectives

The primary objective is to design a system that can detect and classify ISL gestures with the help of computer vision and deep learning. Once identified, gestures are transformed into readable text and optionally converted into speech using text-to-speech (TTS) processing. A secondary objective is to develop a well-structured dataset based on official ISL standards, enabling consistent and accurate training for the recognition model. The broader aim is to build a reliable, accessible solution that enhances communication between signers and non-signers.

II. LITERATURE REVIEW

Earlier contributions, such as those by Shenoy et al. (2018) [1], utilized classical image-processing techniques in combination with machine learning algorithms to recognize both static and dynamic ISL gestures, forming a foundation for later deep learning-oriented solutions.

Seeking further improvements, Murali et al. (2020) [2] developed a CNN-based framework capable of efficiently handling gesture image sequences. Research on sign language recognition has advanced substantially over the past decade, shifting from basic image-processing methods to sophisticated deep learning frameworks.

In a related effort, Mohammedali et al. (2022) [3] designed a computer-vision-driven hand-tracking pipeline that highlighted the significance of precise feature extraction and gesture segmentation, achieving reliable interpretation of various sign patterns.

Building upon these developments, Puranik et al. (2022) [4] employed Recurrent Neural Networks (RNNs) to analyze continuous video streams, effectively modeling the temporal characteristics required for interpreting dynamic sign movements.

Goyal (2023) [5] proposed an ISL recognition system based on MediaPipe Holistic, using its integrated hand, face, and body landmark extraction features to streamline preprocessing and enhance real-time gesture detection.

III. METHODOLOGY

The development process for the ISL translation system consists of several key stages, each contributing to accurate gesture recognition and smooth real-time interaction.

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

3.1 Data Collection

A custom dataset of ISL alphabet gestures (A–Z) is compiled using images captured under varied lighting and background conditions. Collecting diverse samples helps improve the robustness of the model and prepares it for real-world scenarios.

3.2 Data Preprocessing

Images are standardized through resizing and normalization to suit the input requirements of the CNN model. The dataset is split into training and testing sets. Augmentation techniques such as rotation, flipping, zooming, and brightness changes are applied to increase dataset variability and reduce overfitting.

3.3 Feature Extraction

CNN layers are used to extract meaningful features from the input images. These layers learn visual cues such as finger shapes, boundary edges, and hand contours, which are essential for distinguishing between similar signs.

3.4 Model Training

The extracted features are passed through fully connected layers that classify each gesture into its corresponding alphabet. Training is carried out using categorical cross-entropy loss and optimized with Adam to ensure stable and efficient learning. Multiple training epochs help the model generalize to new, unseen gestures.

3.5 Real-Time Recognition

After training, the model is integrated with OpenCV and MediaPipe. A webcam feed provides real-time input, MediaPipe extracts hand landmarks, and the trained CNN predicts the associated alphabet. The system delivers predictions at interactive frame rates, ensuring a fluid user experience.

3.6 Text and Speech Output

Recognized gestures are displayed on a user interface in textual form. To enhance accessibility, the system additionally incorporates TTS technology, converting the detected text into speech to facilitate communication between signers and non-signers.

IV. SYSTEM ARCHITECTURE

The architecture of the proposed ISL recognition system is organized into four sequential processing stages, each responsible for a specific function within the translation pipeline. Figure 4.1 shows the system architecture. The system begins with the Input and Preprocessing phase, where hand gestures are captured through a camera interface. The acquired frames undergo essential preprocessing operations such as normalization, hand-region isolation, and noise reduction to ensure uniformity and improve downstream model performance.

Figure 4.1: ISL text/speech system architecture

ISSN (Online): 2583-648X

IJIRID

International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

In the Feature Extraction stage, the preprocessed images are analyzed to derive meaningful descriptors that represent the structure and orientation of the hand. Techniques such as geometric landmark detection or deep feature extraction are employed to capture critical characteristics of the gesture while reducing irrelevant information.

The extracted features are then fed into the Recognition and Classification module. This component utilizes a trained machine learning or deep learning model to match the incoming feature vectors with learned patterns corresponding to ISL alphabets or gesture classes. The classifier outputs the most probable gesture based on the model's predictive capability. Finally, the Translation and Output stage converts the recognized gesture into an accessible communication form. The predicted class label is displayed as text, and when integrated with text-to-speech processing, can also be rendered as spoken output. This final stage enables seamless interaction between individuals using sign language and those who rely on spoken or written communication.

V. RESULTS AND DISCUSSION

The trained CNN model demonstrated strong performance in recognizing ISL alphabet gestures. Accuracy and loss curves for both training and validation indicated stable learning, with no evidence of overfitting. The confusion matrix showed high precision across most classes, with minor ambiguities arising in gestures with similar visual patterns such as U vs. V or M vs. N which closely resemble each other in hand structure.

Screenshot 5.1: Hand Sign Detection (Number Recognition)

In real-time testing, MediaPipe enabled efficient landmark extraction, and the system executed predictions smoothly at live frame rates. The model performed reliably under normal lighting conditions but exhibited slight performance reductions in low-light or highly cluttered environments. Despite these challenges, the system successfully translated static ISL gestures into text with high accuracy, demonstrating its potential as a practical assistive communication tool.

ISSN (Online): 2583-648X

IJIRID

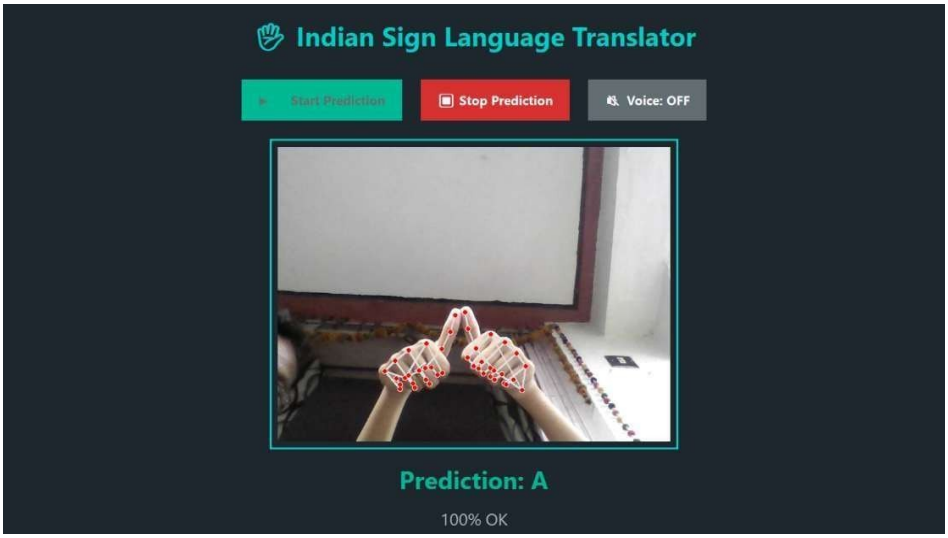
International Journal of Ingenious Research, Invention and Development

An International, High Impact Factor, Double-Blind Peer-Reviewed, Open-Access, Multidisciplinary Online Journal

Volume 4 | Issue 5 | October 2025

www.ijirid.in

Screenshots of hand sign detection for numbers and alphabets are shown below to illustrate system functionality.



Screenshot 5.2: Hand Sign Detection (Alphabet Recognition)

VI. CONCLUSION

This project presents a fully functional ISL recognition system capable of translating alphabet gestures into text and speech using a CNN model combined with MediaPipe and OpenCV. The system is accurate, responsive, and cost-effective, offering a software-only alternative to hardware-based gesture recognition solutions. While the current implementation focuses on static alphabet gestures, the framework can easily be extended to dynamic signs, word-level translation, and bidirectional communication in future work. Enhancing the dataset and experimenting with more advanced neural network architectures can further improve overall reliability. Ultimately, the system demonstrates how AI-driven technologies can support greater accessibility and inclusivity for individuals with hearing and speech impairments.

REFERENCES

[1] K. Shenoy et al., "Real-time Indian Sign Language (ISL) Recognition," in Proc. 9th Int. Conf. Computing, Communication and Networking Technologies (ICCCNT), 2018.

[2] R. S. L. Murali, L. D. Ramayya, and V. A. Santosh, "Sign Language Recognition System Using Convolutional Neural Network and Computer Vision," 2020.

[3] A. H. Mohammedali, H. H. Abbas, and H. I. Shahadi, "Real-time Sign Language Recognition System," International Journal of Health Sciences, vol. 6, pp. 10384–10407, 2022.

[4] V. Puranik, V. Gawande, J. Gujarathi, A. Patani, and T. Rane, "Video-based Sign Language Recognition Using Recurrent Neural Networks," in Proc. 2nd Asian Conf. Innovation in Technology (ASIANCON), IEEE, Aug. 2022, pp. 1–6.

[5] K. Goyal, "Indian Sign Language Recognition Using Mediapipe Holistic," arXiv preprint arXiv:2304.10256, 2023.

