© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712873

# Indian Sign Language Recognition System

A. R. PRAGNA[1], AKKI LIKITHA[2], HEMA SREE[3], KAVYA K[4], SHEELA B. P[5], DR. B. SREEPATHI[6]
[1, 2, 3, 4, 5, 6]*Department of ISE, RYMEC, Ballari*

***Abstract- Communication barriers between the hearing-impaired, speech-impaired, and non-signers create persistent challenges in education, healthcare, and daily interactions. While Indian Sign Language (ISL) is widely used within the deaf community, the lack of public awareness significantly limits effective communication. This paper presents a real-time ISL Recognition System developed using deep learning and computer vision. The system captures hand gestures through a webcam, extracts 21 hand landmarks using Google Mediapipe, and classifies them using a trained Convolutional Neural Network (CNN). The recognized gestures are translated into text and speech for better accessibility. In addition to gesture recognition, the system provides speech-to-text functionality to enable two-way communication between hearing and non-hearing users. The proposed solution is lightweight, cost-effective, and easy to deploy, making it suitable for educational institutions, assistive technologies, and inclusive digital platforms. The results demonstrate high accuracy and robust performance, validating the effectiveness of the proposed framework.***

***Keywords: Indian Sign Language, Deep Learning, Gesture Recognition, CNN, Mediapipe.***

## I. INTRODUCTION

Communication is an essential part of human interaction, enabling individuals to share ideas, express emotions, and participate in society. However, millions of individuals who are hearing-impaired or speech-impaired face significant challenges while interacting with people who do not understand Indian Sign Language (ISL). Although ISL is standardized and widely used across India, it remains unfamiliar to the general population, leading to misunderstandings and exclusion in social, professional, and academic environments.

Advancements in Artificial Intelligence (AI), particularly in deep learning and computer vision, have opened the door for systems that can automatically interpret human gestures. Traditional systems for sign language recognition often rely on hardware-based sensors, specialized gloves, or depth cameras, which are expensive and impractical for everyday use. With the emergence of vision-based frameworks like Mediapipe and TensorFlow, it has become possible to develop low-cost, real-time ISL recognition systems that require only a standard webcam.

The main objective of this work is to design a system capable of recognizing ISL gestures and converting them into meaningful text and speech. This system not only supports gesture-to-speech but also integrates speech-to-text functionality, allowing two-way communication. The proposed approach is simple, scalable, and suitable for deployment in classrooms, public service centers, and assistive technology applications.

## II. PROBLEM STATEMENT

Despite the benefits of sign language, the lack of universal understanding creates communication gaps that isolate hearing-impaired individuals. Existing recognition systems are either expensive, limited in accuracy, or dependent on controlled environments. Many systems also support only one-way communication — converting gestures to text — without enabling the reverse flow of information.

There is a need for a real-time, camera-based ISL recognition system that works in natural environments, eliminates the requirement for specialized hardware, and enables smooth two-way interaction. The proposed system addresses these gaps by integrating gesture recognition and speech processing in a single unified framework.

## III. OBJECTIVES

To address the challenges mentioned, the system is designed with the following objectives:

- To develop a cost-effective, webcam-based sign language recognition system.
- To apply machine learning and deep learning methodologies for robust gesture classification.
- To implement a CNN model capable of

© DEC 2025 | IRE Journals | Volume 9 Issue 6 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I6-1712873

extracting unique spatial features from hand landmarks.

- To provide text and speech output for recognized gestures, improving accessibility.
- To integrate speech-to-text conversion to enable two-way communication.
- To design a simple, intuitive, and user-friendly interface for differently-abled users.
- To ensure real-time performance with minimal computational requirements.

## IV. EXISTING SYSTEM

Existing ISL recognition systems suffer from multiple limitations. Sensor-based systems use gloves equipped with flex sensors or accelerometers to detect hand movements. While accurate, these devices are costly and uncomfortable for long-term use. Camera-based systems, on the other hand, often rely on background removal, controlled lighting, or pre-recorded datasets, making them impractical for real-world deployment.

Many academic works focus on static datasets and fail to address real-time detection challenges such as hand occlusion, variations in illumination, gesture overlapping, and diverse hand shapes. Moreover, current systems typically lack speech integration, making them unsuitable for two-way communication. These limitations motivated the development of a robust, lightweight, vision-based ISL recognition system.

## V. PROPOSED SYSTEM

The proposed system uses a webcam to capture ISL gestures and employs Mediapipe to extract 21-keypoint hand landmarks in real time. These landmarks form the input feature vector for a CNN model that identifies the gesture with high accuracy. The prediction is displayed on the screen and converted into speech using text-to-speech technology, enhancing accessibility for non-signers.

To ensure inclusiveness, the system also supports speech-to-text conversion, enabling a bidirectional communication loop. Users can speak into the microphone, and the system translates the speech to text, allowing hearing-impaired individuals to understand spoken language without requiring lip-reading or external support.

The system's architecture is modular and extensible, allowing additional gestures and dynamic sign sequences to be added in the future.

## VI. METHODOLOGY

### 6.1 Data Design

The dataset consists of hand landmark coordinates captured using Google Mediapipe. Each gesture sample contains 21 landmarks, with each landmark defined by its (x, y, z) coordinates. These features are stored in a CSV file and pre-processed before training. The dataset includes multiple samples of each gesture to improve the robustness and generalization ability of the model.

### 6.2 System Workflow

The raw landmark data is normalized to ensure scale invariance across different users. Noise filtering and smoothing techniques are applied to handle slight hand tremors. The dataset is also balanced to prevent bias toward certain gestures during training.

### 6.3 System Architecture

- Image Capture: The webcam streams live video.
- Hand Detection: Mediapipe detects the hand region and extracts21 landmarks.
- Feature Extraction: Landmark coordinates are flattened and normalized.
- Gesture Classification: A CNN model trained on the dataset predicts the gesture.
- Output Generation: The recognized gesture is displayed as text and converted into speech.
- Speech Input: The system accepts microphone input and converts it into text.

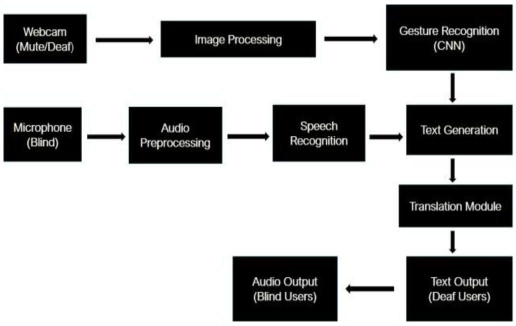



Fig :- Architectural Design

6.4 Model Training

The CNN model is trained using a dataset of labeled gestures. It uses multiple convolutional layers to extract spatial relationships between landmarks, followed by dense layers for classification. The model is optimized using Adam optimizer and categorical cross-entropy loss. The training achieved high accuracy with minimal overfitting due to proper use of dropout and batch normalization.

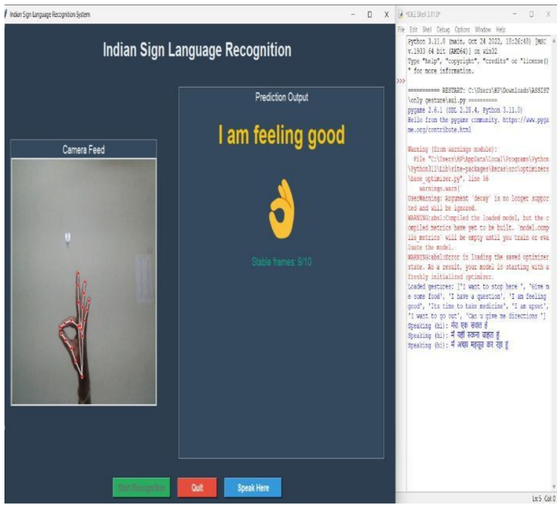
6.5 User Interface Design

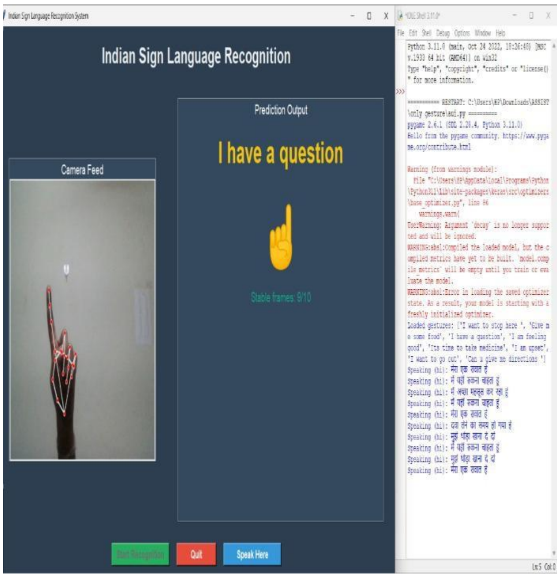
The interface provides real-time camera preview, gesture prediction text, and speech buttons. It is designed for ease of access, enabling disabled users to operate it without difficulty.

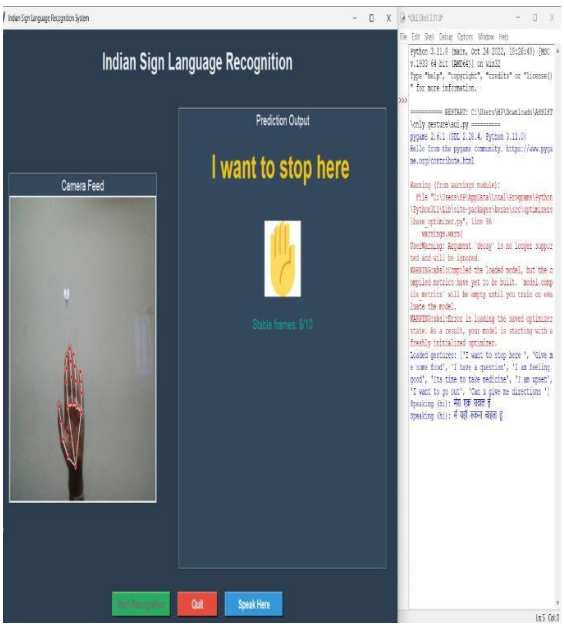
## VII. RESULTS

The system was tested under various lighting conditions and backgrounds to evaluate its robustness. Experiments showed that the model achieved an accuracy of approximately 92–95% for static ISL gestures. Real-time performance was smooth, with minimal delay between hand movement and model prediction. The speech output was clear and understandable, making the system suitable for practical communication scenarios.

The speech-to-text module performed well for most commonly used English words and sentences. User testing indicated that the interface was intuitive and simple to navigate, even for individuals with limited technical knowledge. The system shows potential for use in classrooms, healthcare centers, and public service environments.




## VIII.CONCLUSION

The Indian Sign Language Recognition System successfully demonstrates how artificial intelligence and computer vision can bridge communication gaps between hearing-impaired individuals and the general population. By combining gesture-to-text, gesture-to-speech,and speech-to-text functionalities, the system creates a complete two-way communication platform. The lightweight and modular design enables easy deployment on standard computers without requiring expensive hardware.

Future enhancements may include dynamic gesture recognition, mobile application development, multi-language speech support, and integration with cloud-based models for improved performance. Overall, the system shows strong potential to support inclusive communication and improve accessibility for differently-abled communities.

## REFERENCES

[1] TensorFlow Documentation, TensorFlow Machine Learning Framework. Available at: https://www.tensorflow.org
[2] Keras API Reference, Deep Learning Library for Python. Available at: https://keras.io
[3] Mediapipe by Google, Cross-platform Machine Learning Framework. Available at: https://mediapipe.dev
[4] OpenCV-Python Tutorials, Open Source Computer Vision Library Documentation. Available at: https://docs.opencv.org
[5] Google Text-to-Speech (gTTS) Documentation, Python Speech Synthesis Library. Available at: https://gtts.readthedocs.io
[6] Research Paper: Real-Time Sign Language Recognition using Deep Learning and Computer Vision, IEEE Publications.
[7] Indian Sign Language Dictionary, Rehabilitation Council of India (RCI). Available at: https://www.indiansignlanguage.org
[8]Brownlee, J., Deep Learning for Computer Vision, Machine Learning Mastery Series, 2020.
[9] S. S. Rautaray & A. Agrawal, Vision-Based Hand Gesture Recognition for HumanComputer Interaction, Springer, 2015.
[10] Python Official Documentation. Available at: https://docs.python.org
[11] Sheela, B.P., Girisha, H. & Sreepathi, B. (2025). *An Efficient EPReLU-CSGNN-MALSTCAM and SSOA-Based Explainable Artificial Intelligence (XAI) to Generate Textual Explanations*. SN Computer Science, 6, 594. https://doi.org/10.1007/s42979-025-04115-w
[12] Sheela, B.P., & Girisha, H. (2024). *An Exploration on Explainable AI with Background and Motivation for XAI*. In: Basha, S.M., Taherdoost, H., Zanchettin, C. (eds) Innovations in Cybersecurity and Data Science (ICICDS 2024). Springer, Singapore. https://doi.org/10.1007/978-981-97-5791-6_36
[13] Sheela, B. P., & Girisha, H. (2024). *An Explainable Artificial Intelligence (XAI) Framework for Deep Learning Based Classification to Generate Textual Explanations on Predicted Images*. International Journal of Intelligent Engineering & Systems, 17(6), 651–662.https://doi.org/10.22266/ijies2024.1231.50
[14] Sheela B. P., Sanganal, K., Girish, Sharanegouda G., Rajashekhar R., & Sreepathi, B. (2023). *Physiotherapy Assistance Application for Physically Disabled People Using Computer Vision*. International Research Journal of Modernization in Engineering Technology and Science (IRJMETS), 5(5), 1256. https://doi.org/10.56726/IRJMETS38402
[15] Sheela B.P., Siddana M., Mounika H., Abhishek N., Swathi N. M., Prof P., & Dr. Sreepathi. (2022). *Crop Yield Prediction based on Indian Agriculture using Machine Learning*. International Journal of Advanced Research in

Science, Communication and Technology, 53–57. https://doi.org/10.48175/IJARSCT-5881