Archives available at journals.mriindia.com

International Journal on Advanced Computer Engineering and Communication Technology

ISSN: 2278-5140

Volume 14 Issue 03s, 2025

AI-Driven Real-Time Sign Language Recognition System

[1]Dr. Reema Roychaudhary, [2]Mrudul Dehankar, [3]Roshan Tigga, [4]Pranjal Kothekar, [5]Nainesh Zod

[1]Assistant Professor Department of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology Nagpur, India
[2,3,4,5] Department of Computer Engineering
St. Vincent Pallotti College of Engineering and Technology Nagpur, India
Email: [1]reemaroychaudhary@stvincentn gp.edu.in, [2]mruduldehankar.22@stvincentn gp.edu.in, [3]roshantigga.22@stvincentngp.ed u.in, [4]pranjalkothekar.22@stvincentng p.edu.in, [5]naineshzod.22@stvincentngp.ed u.in

Peer Review Information

Submission: 05 Nov 2025

Revision: 25 Nov 2025

Acceptance: 17 Dec 2025

Keywords

Abstract

Communication is a fundamental human right essential for social inclusion. Individuals with hearing or speech impairments face barriers due to limited understanding of sign language. This project presents an AI-driven real-time sign language recognition system that translates gestures into readable text or speech. The system recognizes both single-hand and dual-hand gestures, including daily expressions and gesture-based computations. It utilizes both Artificial Intelligence (AI) and Computer Vision for accurate gesture interpretation. OpenCV for image preprocessing, and MediaPipe for hand landmark detection. A Flask-based web interface enables real-time interaction, while the gestures are recognized and converted into text. The system operates efficiently with high accuracy and can interpret words like 'SUN'. This AI-based solution enhances assistive communication and promotes inclusivity in education, healthcare, and accessibility applications

Introduction

Sign language recognition systems are crucial in providing a means of communication between deaf or mute persons and the rest of the world. Communication is a basic human right, but deaf or mute persons have always lived in seclusion and have difficulties in availing themselves of essential services. With the development of Artificial Intelligence and Computer Vision, real-time gesture recognition systems accurately identify the hand gestures and translate them into written text or spoken words. The research proposes a browser-based sign language recognition web platform that can recognize Single-Hand and Dual-Hand gestures. In addition, the system is further extended to perform gesture-based calculations and the most used phrases, enabling practical communications such as conveying the message "I want water" effectively.

Literature Review

With the continuous growth in information technology, forms of interaction that involve man and computers have seen remarkable changes. Modern interfaces are becoming increasingly intuitive, multimodal, and intelligent in this regard, easing man–computer communication as much as possible. Of the many research areas involved within this field, one such domain that has been particularly influential pertains to the facilitation of communication for the hearing- and speech-impaired by way of sign language recognition systems. Quite a significant amount of research effort has been invested in this area of

communication in order to bridge the gap between deaf and able-bodied people by using digital systems effectively. [1]
Since sign language is a structured visual language composed of hand gestures, facial expressions, and body postures, its recognition lies within the scope of Human–Computer Interaction and Computer Vision. Various methodologies have been explored to capture and interpret these gestures accurately. Generally, sign language recognition systems are divided into two categories: data glove-based systems and computer vision-based systems.[2]
The data glove approach involves the use of wearable gloves fitted with electromechanical sensors that capture and digitize hand and finger motions into numerical format. Though this can facilitate accurate tracking, it has several drawbacks in terms of high cost, difficult setup, and loss of naturalness due to its wearable hardware requirement. These limit the real implementation in natural communication scenarios.[3]
The computer vision-based approaches, in contrast, use only a camera to capture the hand gestures of the users, thus making them more natural and non-intrusive, with no external devices attached. These systems make use of enhanced processing of images along with machine learning to identify gestures from raw visual data. For example, Indian researchers created recognition models for ISL using keypoint detection algorithms like SIFT, where the features of the new input images are compared against stored ones in a pre-trained database to classify the alphabetic gestures.[4]
Further enhancements were made by using edge detection and color-based segmentation, which included the application of bilateral filtering on depth images to refine boundaries and improve recognition accuracy. With the rapid development of deep learning, the use of CNNs and RNNs in modern systems enables the learning of spatial and temporal patterns of gestures automatically, hence a significant improvement of accuracy, speed, and robustness. The integration of AI-driven models with computer vision revolutionizes real-time sign language detection, making communication platforms more inclusive, efficient, and intelligent. [5]

Workflow Of The Project

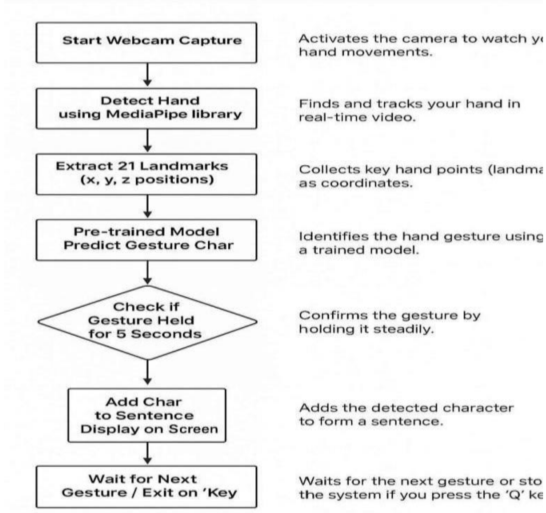

Fig 1: Workflow of the project

Implementation And Proposed Methodology

A] GESTURE COLLECTION (Dataset creation)

1)Identifying Hand Landmarks:

MediaPipe's Hand Tracking solution detects 21 key points (landmarks) on each hand, including fingertips, joints, and the wrist.These landmarks provide precise 3D coordinates for each finger, which are crucial for recognizing the gestures used in sign language.

2) Real-time Gesture Recognition:

By capturing the spatial position and movement of fingers and hands in real time, MediaPipe allows systems to translate gestures into corresponding letters, words, or phrases.This enables continuous sign language detection for live communication.

3) Integration with Machine Learning Models:

MediaPipe acts as a feature extractor. The coordinates of hand landmarks can be fed into ML classifiers (e.g., SVM, Random Forest, LSTM, or CNN) to predict the corresponding sign language symbol.It can also handle temporal sequences to detect signs that involve motion over time.

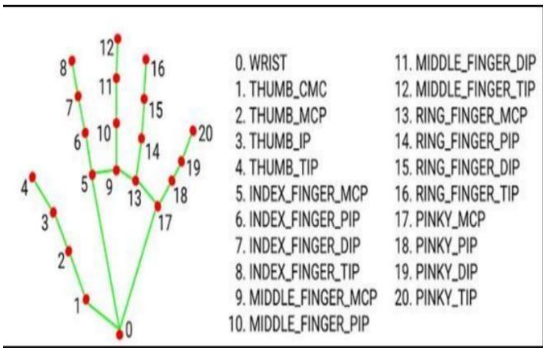

Fig. 2: Detected 20 key points on hand marked by MediaPipe

B] CREATE GESTURE:
Saving Gestures Automatically:
Data Capture:
MediaPipe provides the coordinates of each gesture in real time. These can be recorded and stored in a structured format such a

Automatic Labeling:
If you assign a gesture name or label before recording, MediaPipe can automatically tag the captured coordinates with that label.

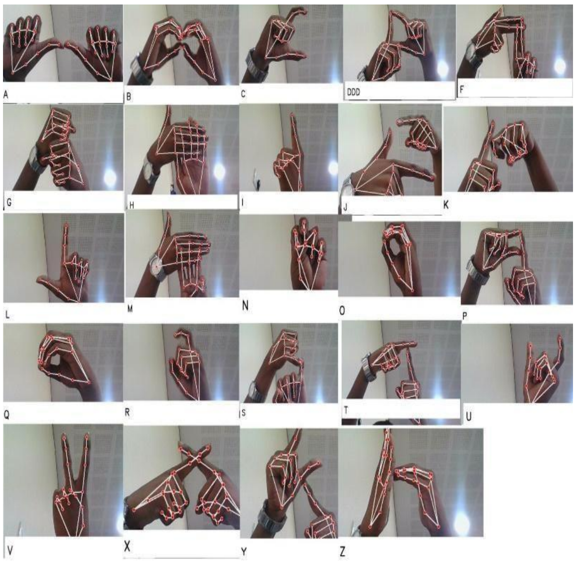

Fig.3: Double Hand Dataset

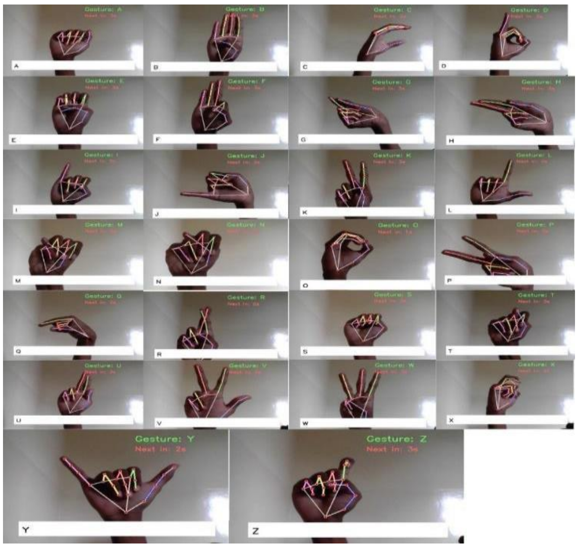

Fig.4: Single Hand Dataset

Technologies Used

The proposed system integrates numerous modern technologies and frameworks that will enable efficient, real-time, and accurate recognition of sign language gestures. Each component plays a specific role in the overall pipeline, ensuring a smooth flow of data acquisition, processing, and output visualization.

A. MediaPipe:

MediaPipe is a technology developed by Google that provides the core functionalities for real-time landmark detection of hands. In fact, it uses a downloaded, pre-trained pipeline to detect up to 21 hand landmarks three-dimensionally through a normal webcam. This is quite efficient because detecting hand movements without the involvement of special sensors or gloves makes it definitely fit for the actual recognition of sign language. Robustness in background variation and lighting changes enables MediaPipe to perform well under all sorts of environments.

B. OpenCV:

Image acquisition, preprocessing, and feature extraction utilize the Open Source Computer Vision Library, OpenCV. This library allows for key operations that improve the quality of gesture images, including resizing, background subtraction, noise reduction, and isolation of regions of interest. With OpenCV integrated into Python, this guarantees speed and efficiency in handling live video streams and lays the foundation for accuracy in gesture detection.

C. Random Forest Classifier:

Gesture classification is done using a Random Forest algorithm. It is an ensemble learning method that constructs multiple decision trees and outputs the most frequent class among them. Random forests are computationally efficient, resistant to overfitting, and perform well on multidimensional datasets like hand landmark coordinates. That makes them appropriate for real-time gesture prediction with high accuracy.

D. Flask Framework:

The trained model is then deployed to a browser-accessible application using Flask, which is a lightweight Python web framework. It provides flawless integration of the backend gesture recognition logic and the front-end user interface. Flask will also enable real-time communication between the user and the system to instantly display recognized gestures as text.

E. Python Environment:

Python acts as the core programming environment for the system's implementation. Its simplicity, flexibility, and extensive support of libraries allow the effective integration of MediaPipe, OpenCV, Flask, and machine learning modules. These technologies together make this AI-driven sign language recognition system accessible, robust, and real-time.

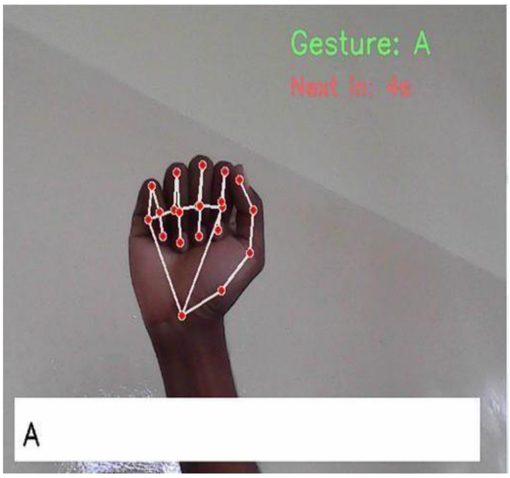

Fig.5: Single Hand Gesture Translation to Text.

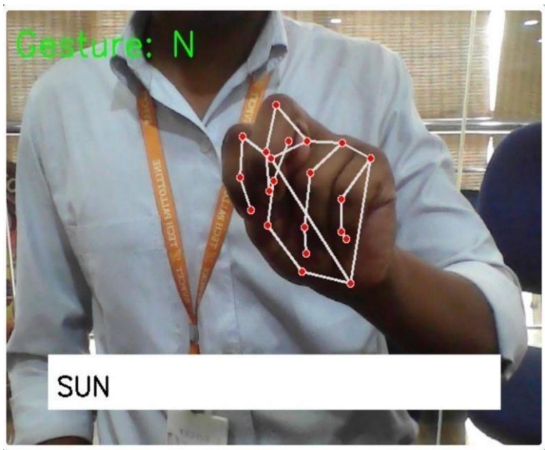

Fig.6: Double Hand Gesture Translation to Text

Result

In this system, a dataset containing all the gestures are present. Each gesture contains approximately 1000-1500 images which is used for training and testing the model. The images are saved automatically according to the landmarks using mediapipe.

The below figures shows the files where the dataset is stored and their coordinates.

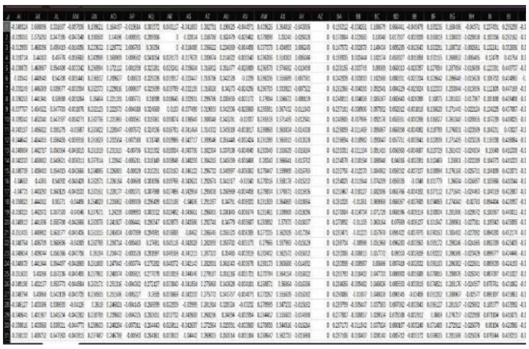

Fig 7: Dataset saved as CSV containing coordinates of hand landmarks

In The below figure the Single and Double Hand Gesture is Translated into the text

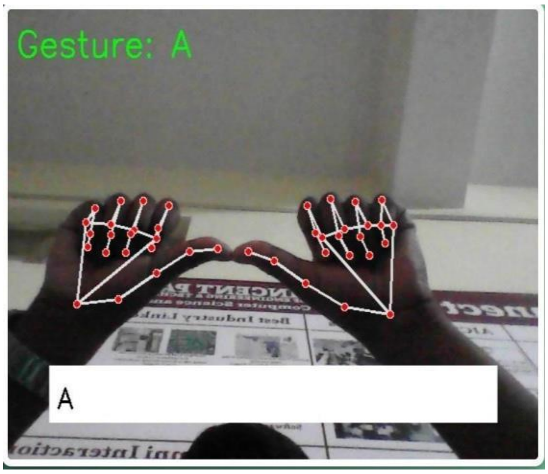

Fig 8: Word using gestures

This bar chart represents the distribution of gesture samples collected for model training. Each bar indicates the total number of samples for a specific gesture label. The dataset is approximately balanced, ensuring fair training for each gesture. Balanced data distribution is crucial for achieving high recognition accuracy and avoiding model bias.

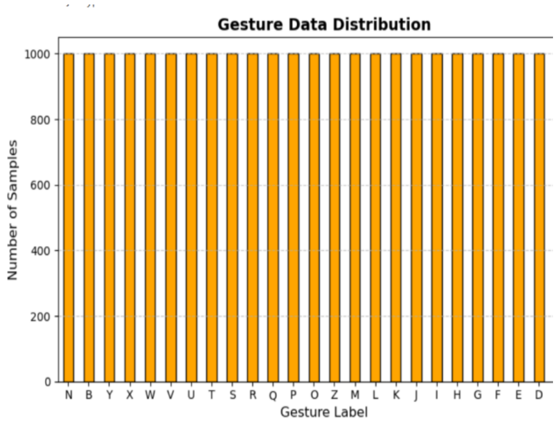

Fig 9: Gesture Data Distribution Chart

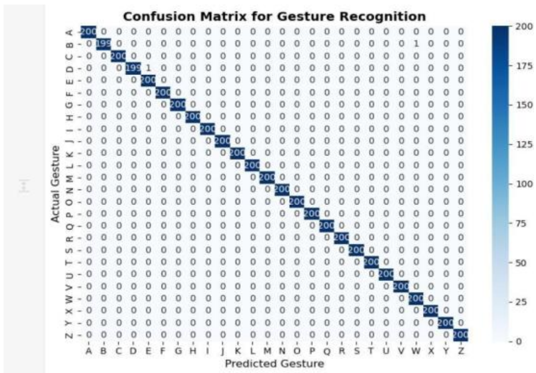

Fig 10: Confusion Matrix for Gesture Recognition

This heatmap visualizes the classification accuracy of the trained RandomForest model for sign language gestures. Each row represents the actual gesture, while each column represents the predicted gesture. The diagonal elements indicate correctly predicted gestures, whereas the off-diagonal values represent misclassifications. A stronger blue color on the diagonal signifies higher accuracy, demonstrating that the model can reliably distinguish between different gestures.

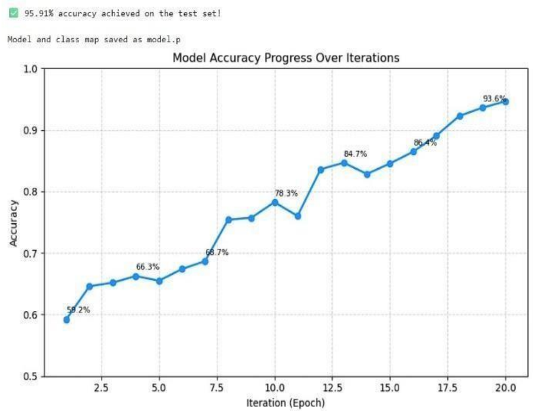

Fig 11 : Graph showing accuracy over iterations(epochs)

Accuracy increases steadily from around **59% to 93.6%**, indicating effective learning and model optimization.
The final **test accuracy achieved is 95.91%**, demonstrating that the system performs with high reliability in recognizing sign language gestures.

Conclusion

The proposed real-time sign language recognition system, driven by AI, is quite effective in bridging the communication gap of hearing- impaired individuals. On the other hand, unlike prior works, for example, Siddharth S., Jayashree R., and Shwetha K. N. (2022), who employed ResNet-50 CNN to recognize Indian Sign Language, reaching an accuracy of ~99.19%, or the work done by Mohamed M. et al. (2021), which made a service-oriented approach towards recognition, our system performs with high accuracy sans CNNs or heavy deep learning frameworks. In our method, Artificial Intelligence, Computer Vision, OpenCV, MediaPipe, and a Flask- based web interface have been employed, which enable the system to perform efficient, real-time recognition of single-hand and dual- hand gestures and common daily expressions. This makes the approach less computationally complex while retaining robust performance. Future enhancements may include expanding the dataset of various gestures, enhancing recognition in different light conditions and complex backgrounds, and employing multimodal features to further improve accuracy and usability. Overall, this project contributes to practical assistive technology, social inclusion, and allows effective communication among hearing- impaired users.

Limitations

Currently, the system can only detect unchanging gestures and is not able to read any words that need movement or several hand movements at the same time. This reduces the ability of the system to translate many natural sign language expressions in everyday conversation from signers.

1. The accuracy of this system is highly defined by the conditions of the light in which it is used. Under low levels of light, very bright light, or when a shadow from the user's hand appears, the MediaPipe program may not be able to accurately identify the various landmarks resulting in inaccurate predictions.
2. In addition, the system tends to be less accurate in environments that have many different things going on in the background. For example, if the background has colour or patterning similar to that of the user's hands, the location of the actual landmarks will not be able to be tracked properly impacting the performance of the entire recognition process.
3. The dataset used to develop the model has a very narrow range and is not diverse. The dataset was developed with hand gestures and did not capture the difference between various skin colours or shades, as well as hand shapes, age groups and lighting that are normally found in different environments. The lack of variety in the dataset has reduced the model's ability to be a general purpose model for all users.
4. The system has been created to focus only on hand gestures and does not take into account any facial expressions or body movements that may be important to the grammatical structure of the signed communication. Therefore, communication is limited because of the system's inability to recognize other aspects of a signer.
5. The MediaPipe model can only be used on a hand or person at any one time. If there are more than one hand within the camera angle of view, the MediaPipe model will not be able to provide a clear view of the landmark positions of the hands, and the recognition results will not be correct.

Future Scope

The proposed AI-driven Sign Language Recognition System lays a strong foundation for real-time gesture interpretation; however, there are several opportunities for enhancement and expansion. With advancements in artificial intelligence, computer vision, and human–computer interaction, the system can be transformed into a highly advanced communication platform. The following points represent an in-depth future scope for the project:

1. Support for Dynamic and Continuous Gestures

Currently, the system recognizes static hand signs. In the future, it can be extended to include:

Dynamic gestures such as waving, movement-based signs, and hand transitions.Continuous sentence recognition, where the system can interpret full-length sign sequences in real time. Integration of time-series models (LSTM, RNN, Transformers) to analyze gesture flow.

2. Voice Output and Speech Synthesis

To make communication more natural, the system can integrate:

Text-to-Speech (TTS) modules to convert recognized gestures into spoken output.Multi-language speech output so users can communicate with anyone, regardless of language barriers.This improvement will significantly help in public places like hospitals, police stations, and schools.

3. Multi-Language Detection and Translation

Future versions can support:

Gesture-to-text conversion in languages such as Hindi,

Marathi, N Tamil, French, etc. Real-time translation of sign language into different regional or international languages.This will broaden the system's usability globally.

4. Development of a Mobile Application

A mobile app (Android/iOS) can make the system more portable and accessible. Using on-device ML, the system could:Work offline without internet.Use smartphone cameras for gesture detection. Assist users in daily interactions such as shopping, traveling, or medical emergencies.

5. Integration of Deep Learning and Advanced Models

While the current system uses Random Forest due to real-time performance, future enhancements may incorporate: Convolutional Neural Networks (CNNs) for richer features. MediaPipe combined with attention-based models for higher accuracy.Hybrid models combining image data + landmarks. This will improve performance for complex signs, variations in lighting, and diverse hand shapes.

6. Larger Dataset Collection

Expanding the dataset will significantly improve robustness. The system can be trained with:

Different skin tones, Multiple age groups,Various lighting conditions,Different hand sizes,Background variations. A diverse dataset will reduce bias and ensure high accuracy in real-world scenarios.

7. Facial Expression and Body Pose Integration

Sign language includes facial cues, head tilts, and body posture. Future versions can include:

Facial expression detection (eyebrows, mouth movements) Upper-body pose estimation .Emotion recognition for more accurate sentence interpretation.

This transforms the system from gesture detection into complete sign language understanding.

8. AR/VR-Based Learning Tools

Augmented Reality and Virtual Reality can be used to:

Teach sign language interactively.Create virtual classrooms for hearing-impaired students. Provide gesture-based training simulations.It will enhance accessibility and learning quality.

9. Deployment in Public Service Infrastructure

The system can be integrated into, Hospitals and clinics for smoother patient communication. Airports, railway stations, and public offices. Banks, courts, and educational institutions.Touch- free communication kiosks can help speech- and hearing-impaired individuals interact easily.

10. Internet of Things (IoT) Integration

By adding gesture control features, the system can allow:

Smart home automation (lights, TV, appliances) Assistive devices for disabled users Gesture-based navigation for computers and smartphones This opens opportunities for gesture-controlled environments.

11. Cloud Deployment and Remote Access

Future versions can be deployed on cloud platforms to enable:

Real-time remote interpretation for video calls. API services for integrating sign recognition into third-party applications. High- performance processing via cloud GPUs.

12. Multi-Person Gesture Recognition

Extend the system to detect and classify gestures from:

Multiple users simultaneously Group environments such as classrooms or meetings This contributes to multi-user communication platforms.

13. Emotion and Context Understanding Adding emotion recognition helps interpret:

Tone of conversation ,User's emotional state, Contextual meaning of certain signs , This will make the system more empathetic and intelligent.

References

Siddharth S., Jayashree R., and Shwetha K. N., "Indian Sign Language Recognition System (2024)," IEEE Conf. on Computational Intelligence and Computing Research, 2022. Utilized ResNet-50 CNN for realtime ISL recognition with

~99.19% accuracy; future work includes model optimization and multimodal features.

Mohamed M., Elnamla M. E., Khamis N. H. H., and Nor Hisham N. A. B., "Sign Language Recognition for ServiceOriented Environments(2024)," IEEE Conf. on Smart Tech for Next Gen Computing, 2021. Used YOLOv5 and CNN for fast ASL detection in real-time settings.

Debnath J., and Praveen Joe I. R., "Real-Time GestureBased Sign Language Recognition(2024)," IJIRCCE, vol. 9, no. 6, pp. 4578– 4585, 2021. Combined OpenCV, MediaPipe Holistic, and LSTM for accurate real-time ASL recognition.

Gupta, S., and Jha, S., "Deep Learning-Based Indian Sign Language Recognition Using CNN-LSTM Model," Proceedings of the 2021 International Conference on Artificial Intelligence and Machine Learning (AIML), IEEE.

Zhang, Z., Li, J., and Hu, W., "Sign Language Recognition UsingConvolutional Neural Networks and Transfer Learning," IEEE Access, vol. 9, pp. 100-112, 2021.

Chen, H., Zhang, Y., and Wang, X., "A Hybrid Deep Learning Model for Sign Language Recognition," International Journal of Pattern Recognition and Artificial Intelligence, vol. 35, no. 8, 2023.

R. Tigga, N. Zod, M. Dehankar, P. Kothekar, and R. Roychaudhary, "AI-Driven Real-Time Sign Language Recognition System," International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 9, no. 8, Aug. 2025.

E.Padmalatha, S.Sailekya, R.Ravinder Reddy, Ch.Anil Krishna ,K.Divyarsha International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Aman Pathak| Avinash Kumar|Priyam|Priyanshu Gupta|Gunjan Chugh International Journal for Modern Trends in Science and Technology, 8(01): 32-37, 2022 International Journal for Modern Trends in Science and Technology

N. RATHIPRIYA AND N. MAHESWARI School of Computer Science and Engineering, Vellore Institute of Technology, Chennai600127, India Corresponding author: N. Maheshwari.