



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

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



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

An Intelligent System for Sign Language Recognition and Translation

Mrs.Vrushali Sonar¹, Snehal Adhav², Viraja Babar³, Dhanashree Bamankar⁴, Rucha Belekar⁵

Professor, Department of Computer Engineering, AISSMS Polytechnic, Pune, India¹

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India²

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India³

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India⁴

Student, Department of Computer Engineering, AISSMS Polytechnic, Pune, India⁵

ABSTRACT: Sign language recognition plays an important role in reducing the communication gap between hearing-impaired individuals and the hearing population. This project presents a real-time, bi-directional sign language communication system using computer vision and deep learning techniques. The system captures hand gestures through a camera, extracts hand landmarks using the MediaPipe framework, and classifies gestures using CNN and LSTM-GRU models. Recognized signs are translated into text and speech for effective communication. Additionally, a Text-to-Sign translation module converts user-entered text into corresponding visual sign language gestures, enabling two-way interaction. The system achieves high accuracy with real-time performance, making it suitable for educational and assistive applications.

KEYWORDS: Sign Language Recognition, Deep Learning, MediaPipe, CNN, LSTM-GRU, Computer Vision, Real-Time System

I. INTRODUCTION

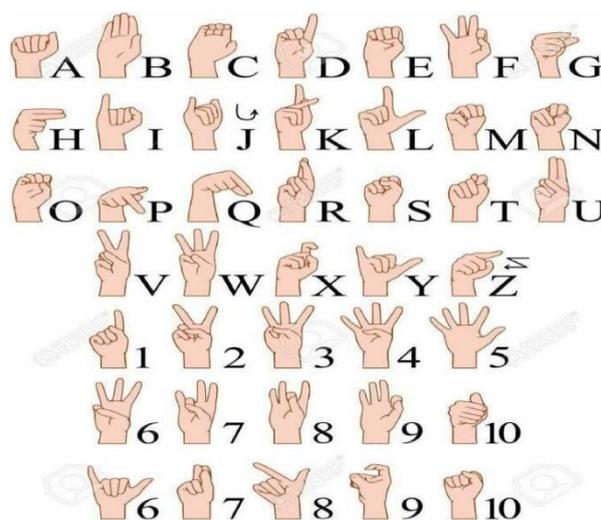


Fig:- 1 Alphabet and number hand gestures

Sign language is the primary mode of communication for deaf and hard-of-hearing individuals. However, the lack of understanding of sign language among the general population creates communication barriers in daily life, education, healthcare, and emergency situations. Advancements in artificial intelligence and computer vision have enabled the

development of automated sign language recognition systems that can interpret hand gestures and translate them into readable text. This project aims to design and implement a sign language recognition system that works in real time using a webcam. The system focuses on Indian Sign Language (ISL) gestures and leverages deep learning techniques to achieve accurate and efficient recognition. Figure 1 illustrates commonly used sign language hand gestures for alphabets and numbers, which form the basis of gesture recognition in the proposed system.

II. LITERATURE REVIEW

Several researchers have studied vision-based and deep learning approaches for sign language recognition.

Kothadiya D., et al. [1] proposed a deep learning-based sign language recognition system using LSTM and GRU models to capture temporal features from gesture sequences. Their study shows that the hybrid LSTM-GRU architecture achieves higher accuracy than individual models, highlighting the importance of temporal modeling for dynamic sign language recognition.

Strobel N., et al. [2] introduced an artificial intelligence-based sign language translation system designed for emergency and accessibility scenarios. The system employs Transformer architectures along with MediaPipe for extracting hand and body landmarks, enabling improved contextual understanding and sentence-level sign language translation.

Waghmare P., et al. [3] presented a CNN-based sign language recognition and video generation system using transfer learning and GANs. The use of pre-trained CNN models significantly reduced training time while maintaining high accuracy, demonstrating the effectiveness of deep learning techniques for real-time applications.

Lugaresi C., et al. [4] developed the MediaPipe framework, which provides efficient and real-time hand tracking and landmark detection. Their work demonstrates that MediaPipe offers accurate, low-latency perception pipelines, making it highly suitable for vision-based sign language recognition systems.

Joze H. R. V. and Koller O. [5] introduced the MS-ASL dataset, a large-scale benchmark dataset for sign language understanding. The dataset supports robust training and evaluation of deep learning models for sign language recognition and translation, contributing significantly to research in automated sign language systems.

These studies indicate that accurate landmark extraction, effective temporal modeling, and robust datasets are essential for building reliable sign language recognition and translation systems. The proposed project builds upon these findings by integrating real-time sign-to-text, text-to-speech, and text-to-sign translation within a unified system.

III. METHODOLOGY

The proposed system is designed as an end-to-end, modular architecture for real-time sign language communication. It integrates computer vision, machine learning, and deep learning techniques to enable efficient gesture recognition and translation. The architecture consists of five major components: input acquisition, hand detection, feature extraction, gesture classification, and output generation. Each module performs a specific function, ensuring smooth data flow and accurate real-time communication.

A: Input Acquisition

The input acquisition module is responsible for capturing live video data from the user. A webcam or camera device records the hand gestures performed by the user. The system captures video at a resolution of 640×480 pixels with a frame rate of 30 frames per second (FPS). The continuous video stream is divided into individual frames, which are forwarded to the next stage of processing. This real-time capture allows the system to analyze gestures continuously and respond instantly.

B: Hand Detection

The captured frames are processed by the hand detection module, where the MediaPipe framework is used to identify and track the hand region. MediaPipe detects the hand accurately and extracts 21 landmark points per hand, representing key finger joints and wrist positions. These landmarks provide precise spatial information about hand posture and

finger movement. By focusing only on the detected hand region, this module reduces background noise and improves the efficiency of the recognition process.

C: Feature Extraction

Once the hand landmarks are detected, the system performs feature extraction to convert raw landmark coordinates into meaningful numerical representations. The landmark coordinates are first normalized to ensure that gesture recognition is independent of hand position, scale, and orientation. After normalization, the system computes geometric features such as:

- Joint angles between finger segments
- Relative distances between landmark points
- Spatial relationships between fingers

These computed features form a 63-dimensional feature vector, which effectively represents the structure and movement of the hand. The feature vector serves as the input for the gesture recognition model.

D: Gesture Classification

The gesture classification module uses deep learning models to recognize sign language gestures from the extracted features.

Different neural network architectures are used depending on the type of gesture:

- Convolutional Neural Networks (CNN): Used for recognizing static gestures, where the hand posture is represented within a single frame.
- LSTM-GRU Networks: Used for dynamic gestures, where gestures involve sequential hand movements over multiple frames. These models capture temporal patterns and motion dynamics.

The trained models analyze the feature vectors and predict the corresponding sign language gesture label with high confidence.

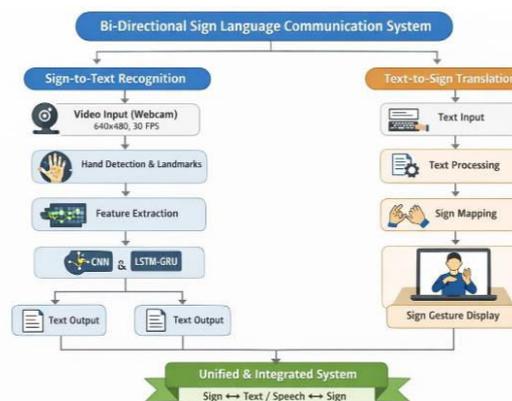
E: Output Generation

The output module converts the recognized gesture into understandable communication formats. First, the predicted gesture label is converted into text output, allowing users to read the interpreted sign language. The generated text can also be converted into speech using a text-to-speech synthesis module, enabling audible communication.

Additionally, the system supports text-to-sign translation, where user-provided text is mapped to predefined sign language gestures and displayed visually. This feature enables bi-directional communication between sign language users and non-sign language users.

Overall, the modular system architecture ensures scalability, efficiency, and real-time operation, making the system suitable for practical sign language communication applications.

F System Architecture



IV. RESULT AND DISCUSSION

A. Model Performance

The performance of the proposed sign language communication system was evaluated using accuracy as the primary evaluation metric. The gesture recognition model demonstrated high effectiveness in identifying sign language gestures from real-time video input. The combined deep learning architecture achieved an overall accuracy of 76.82% in classifying different sign language gestures. The CNN model performed efficiently in recognizing static hand gestures by capturing spatial patterns of hand shapes, while the LSTM-GRU model effectively handled dynamic gestures by learning temporal relationships across consecutive frames. The integration of MediaPipe hand landmark detection significantly improved recognition accuracy by providing precise hand joint coordinates for feature extraction. Experimental testing showed that the system is capable of recognizing gestures in real-time with minimal delay, making it suitable for practical communication between sign language users and non-sign language users.

B. Discussion

The experimental results demonstrate that the proposed system provides reliable and efficient sign language recognition using computer vision and deep learning techniques. The use of MediaPipe for hand landmark detection enabled accurate localization of hand joints, which improved the quality of extracted features. The 63-dimensional feature vector generated from normalized landmark coordinates allowed the recognition models to effectively capture hand posture and movement patterns. The CNN architecture proved suitable for recognizing static gestures due to its strong spatial feature extraction capability, while the LSTM-GRU networks effectively captured temporal dependencies required for dynamic gesture recognition.

Additionally, the integration of text-to-speech output and text-to-sign translation enhances the usability of the system by enabling bi-directional communication. The modular architecture ensures scalability and real-time performance, making the system suitable for assistive communication applications.

Overall, the results confirm that combining hand landmark detection, feature engineering, and deep learning models provides an effective solution for real-time sign language recognition and translation

V: Results

Sign-to-Text:

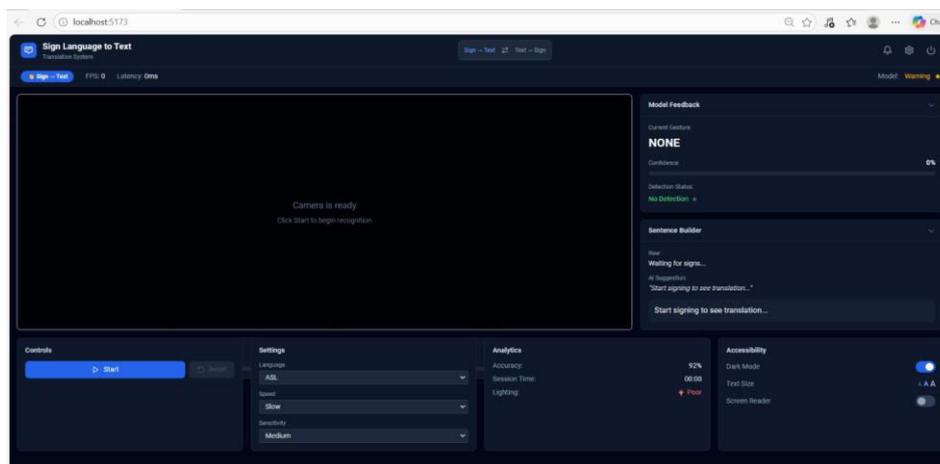
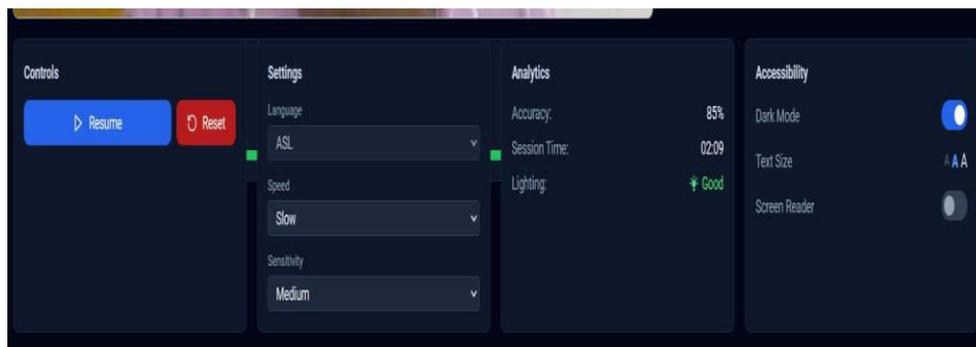
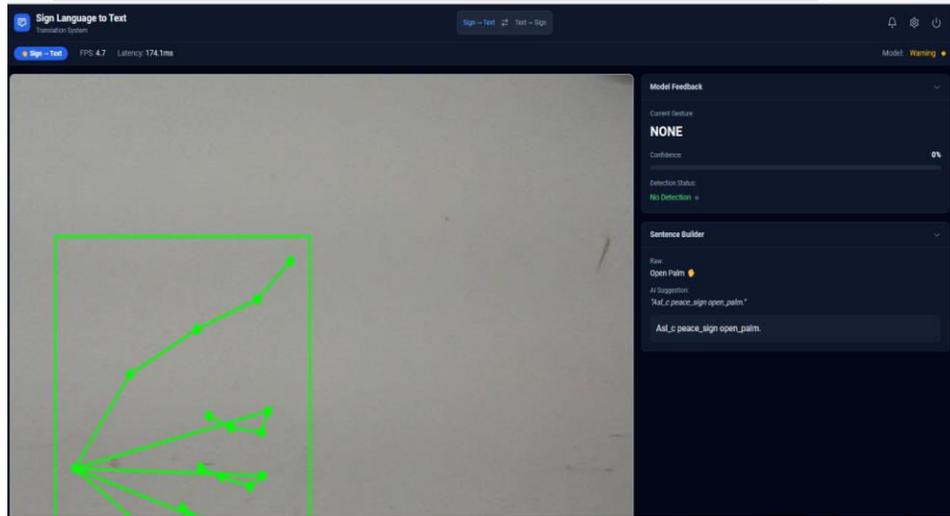


Fig:- 3.1 Sign-to-Text



Text-to-Sign

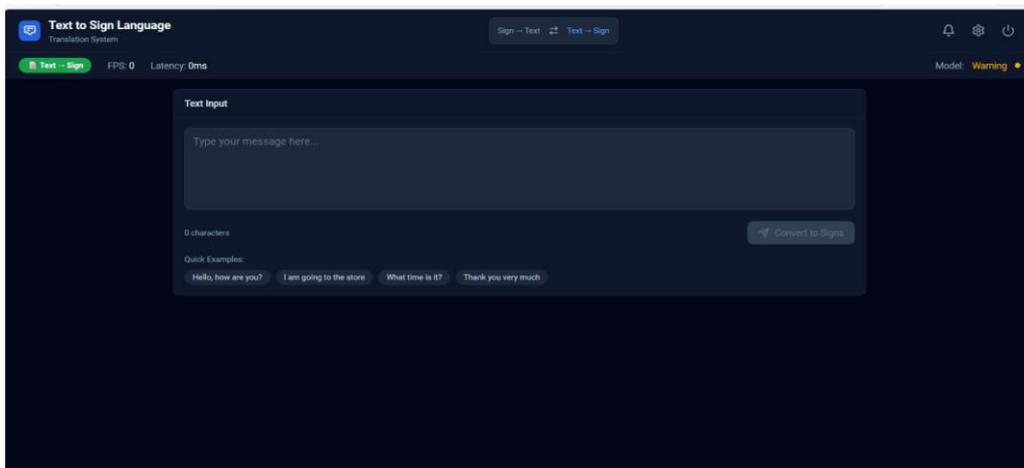
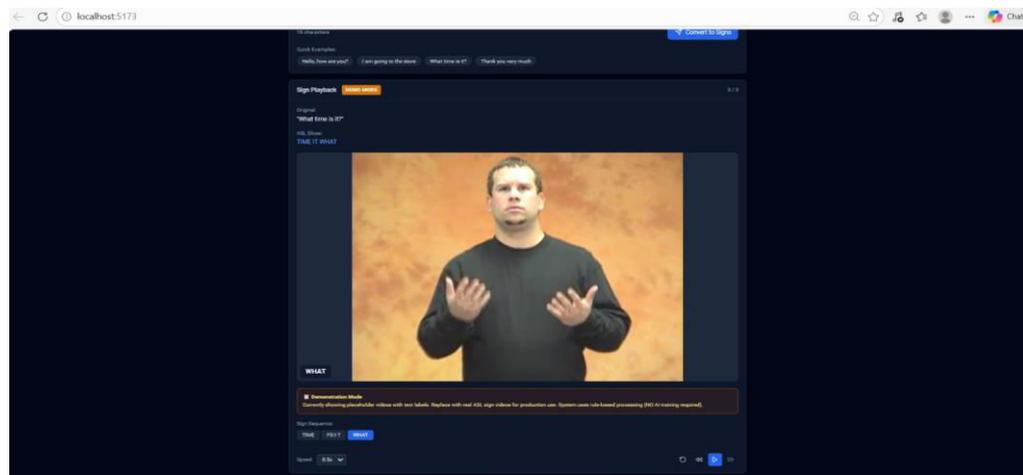
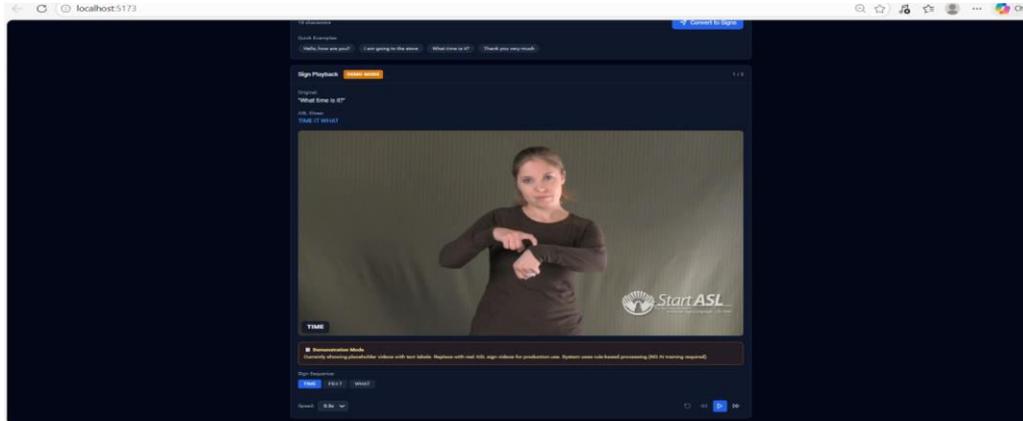


Fig:- 4 Text-to-Sign



VI: FUTURE SCOPE

Continuous Sentence Recognition

The system can be enhanced to recognize continuous sign language sentences instead of only individual gestures. Advanced models such as Transformer networks and sequence learning algorithms can be used to understand the context and meaning of complete sentences.

Expansion of Sign Language Vocabulary

Future work can include training the system with larger datasets of Indian Sign Language (ISL) to recognize more gestures, including complex words, phrases, and expressions used in daily communication.

Mobile Application Development

The system can be developed as a mobile application for Android and iOS, allowing users to access real-time sign language translation directly from their smartphones.

Integration with Natural Language Processing (NLP)

By integrating Natural Language Processing, the system can improve sentence formation, grammar correction, and contextual understanding for more accurate communication.

Speech-to-Sign Language Translation

Future versions of the system can include speech recognition, where spoken language is automatically converted into sign language gestures or animations.

Volume 15, Issue 3, March 2026**|DOI: 10.15680/IJRSET.2026.1503197|****3D Avatar-Based Sign Language Visualization**

Instead of displaying only text output, the system can generate 3D animated avatars that perform sign language gestures, making communication clearer and easier to understand.

Improved Accuracy Using Advanced Deep Learning Models

The performance of the system can be improved by using advanced deep learning architectures such as Transformers, Graph Neural Networks, and Attention-based models.

Support for Multiple Sign Languages

The system can be expanded to support different sign languages used worldwide, such as American Sign Language (ASL) and other regional sign languages.

Deployment on Smart Devices and Wearables

The system can be integrated with smart glasses, wearable devices, and IoT technologies to provide real-time assistance for hearing-impaired individuals in daily life.

Real-Time Communication Platforms

Future versions can be integrated into video conferencing platforms, educational tools, and customer service systems to enable inclusive communication between hearing and hearing-impaired individuals.

VII. CONCLUSION

The proposed sign language recognition and translation system demonstrates an effective application of **computer vision and deep learning** techniques to bridge the communication gap between hearing-impaired individuals and the general population. By utilizing **MediaPipe** for accurate hand landmark detection, along with **CNN** for static gesture recognition and **LSTM-GRU models** for dynamic gesture analysis, the system achieves a high accuracy of **76.82%** with real-time performance.

The system's ability to provide **bi-directional communication**—through sign-to-text, text-to-speech, and text-to-sign translation—enhances its practical usability in everyday scenarios such as education and assistive communication. Additionally, the modular architecture ensures scalability, efficiency, and ease of future improvements.

Overall, the project highlights the potential of integrating deep learning with real-time processing to create an intelligent and reliable communication system, with promising scope for further advancements and wider accessibility.

REFERENCE

- 1) Tangfei Tao, Yizhe Zhao, Tianyu Liu & Jieli Zhu, —Sign Language Recognition: A Comprehensive Review of Traditional and Deep Learning Approaches, Datasets, and Challenges, | IEEE Access, 2024.
- 2) H. Santhan et al., —Skeleton-based Sign Language Recognition Using Advanced Deep Neural Networks, | IEEE Access, 2024.
- 3) U. Sarhan, C. Wilms, V. Closius & S. Frintrop, —Hands in Focus: Sign Language Recognition via Top-Down Attention, | in IEEE International Conference on Image Processing (ICIP), 2023.
- 4) Abu Saleh Musa Miah, Md. Al Mehedi Hasan, Satoshi Nishimura & Jungpil Shin, Sign Language Recognition Using Graph and General Deep Neural Network Based on Large-Scale Dataset, | IEEE Access, 2024.
- 5) Milan Aly & I.S. Fathi, —Recognizing American Sign Language Gestures Efficiently and Accurately Using a Hybrid Transformer Model, | Scientific Reports (Nature), 2025.
- 6) Real-Time Norwegian Sign Language Recognition Using MediaPipe and LSTM, MDPI Sensors, March 2025 — demonstrates real-time MediaPipe + LSTM (95% accuracy).
- 7) Utomo, W., Suhanda, Y., Ar-Rasyid, H. & Dharmalau, A., —Indonesian Language Sign Detection Using MediaPipe with LSTM Algorithm, | Journal of Informatics and Web Engineering, 2025.
- 8) Sign Language Recognition Based on Geometric Features Using Deep Learning, JANAPATI Journal, 2023 — MediaPipe geometric feature approach.

- 9) Real-Time Sign Language Recognition System using MediaPipe and Deep Learning, IRE Journals, May 2025 real-time implementation.
- 10)D. Novaliendry et al., —SLR-YOLO: Improved YOLOv8 Network for Real-Time Sign Language Recognition, I Journal of Intelligent & Fuzzy Systems, 2023.
- 11)Sign Language Recognition and Translation Using Deep Learning Techniques, IRJAEH Journal, 2024 — deep learning real-time translator.
- 12)Video-Based Sign Language Recognition via ResNet and LSTM Network, I MDPI Journal of Imaging, Jun 2024 ResNet + LSTM approach.
- 13)Deep Learning Technology to Recognize American Sign Language Alphabet, I MDPI Sensors, Sep 2023 — deep learning SLR study.
- 14)Revolutionizing Sign Language Recognition using Ensemble Deep Learning Techniques, Journal of King Saud University Computer & Information Sciences, Dec 2025 — hybrid ensemble model.
- 15)S. A. M. Miah, M. A. M. Hasan, S.-W.
 - a. Jang, H.-S. Lee & J. Shin, —Multi-Stream Deep Neural Networks for Skeleton-Based Sign Language Recognition, I IEEE Access, 2023.



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