Impact Factor (QJIF): 8.38 E-ISSN: 2583-7206, P-ISSN: 3049-2432

International Journal of All Subject Research

Received: 07/February/2026 **IJASR:** 2026; 5(2):69-72 **Accepted:** 12/March/2026

AI-Based Real-Time Sign Language Translator to Text and Audio Using Computer Vision and Deep Learning

[1]Dr. K Karuppasamy, [*2]Mothilalnehru M, [3]Bevis Suruthin M, [4]Navin G and [5]Sabari K

[1]Head of the Department, Department of Computer Science and Engineering, RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

[*2, 3, 4, 5]Department of Computer Science and Engineering, RVS College of Engineering and Technology, Coimbatore, Tamil Nadu, India.

Abstract

This paper presents an AI-based real-time sign language translation system that converts hand gestures into text and audio using computer vision and deep learning techniques. The system captures live gestures through a camera, processes them using trained models, and generates readable text and speech output. The proposed solution is cost-effective and easy to deploy without specialized hardware. It can be effectively used in education, healthcare, and public service environments. The system enhances accessibility, promotes inclusivity, and supports independent communication for hearing-impaired individuals.

Keywords: Sign Language Recognition, Computer Vision, Deep Learning, Real-Time Translation, Gesture-to-Speech.

1. Introduction

Communication is essential in daily life, yet deaf and mute individuals often face difficulties due to the limited understanding of sign language. With advancements in Artificial Intelligence, Computer Vision, and Deep Learning, it is now possible to automatically recognize sign language gestures and convert them into text and audio in real time. This AI-based approach helps bridge the communication gap and promotes inclusive and accessible interaction.

2. Objectives

- To recognize sign language gestures in real time using AI-based models.
- To convert identified gestures into meaningful textual information.
- To generate clear audio output from the translated text
- To minimize communication barriers for deaf and mute individuals.
- To design a low-cost, portable, and user-friendly system.
- To apply modern Artificial Intelligence and Deep Learning techniques for accurate gesture recognition.

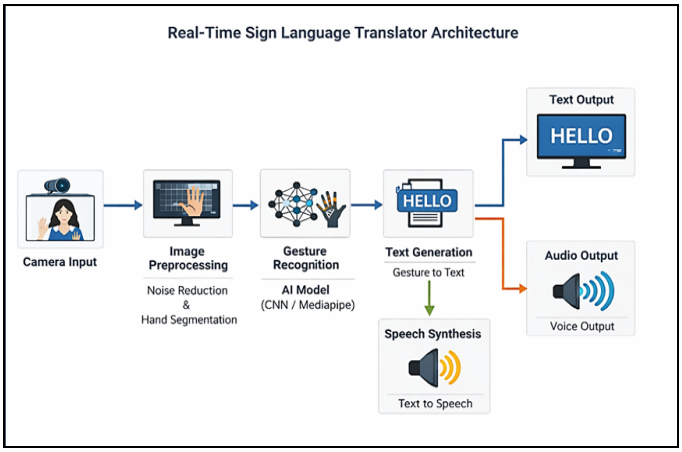

Fig 1: System Architecture of the Real-Time Sign Language Translator

***Corresponding Author:** Mothilalnehru M

3. Existing System

Description

- Communication depends on human interpreters
- Manual sign language translation is required
- Time-consuming and not always available
- High dependency on trained professionals
- Limited accessibility in emergency situations

Limitations:

- Interpreter availability is low
- High cost
- Not scalable
- No real-time automated translation

4. Proposed System

Description:

The proposed system implements a camera-based AI-driven sign language recognition framework that enables automatic translation of hand gestures into text and audio output. The system captures real-time gesture movements and processes them using computer vision and deep learning models to identify meaningful sign patterns. By applying intelligent feature extraction and classification techniques, the system delivers fast and accurate translation without the need for human intervention.

Advantages

- Real-time sign language translation
- Eliminates the need for human interpreters
- High accuracy achieved through deep learning models
- Portable and easily scalable architecture
- Enhances accessibility and social inclusivity

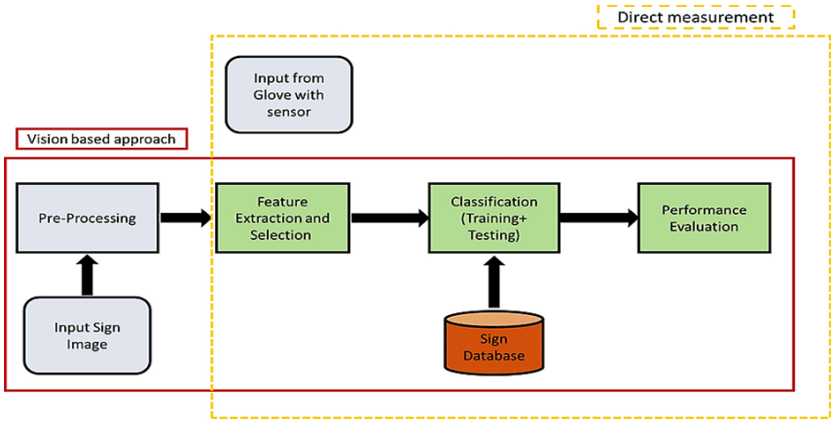

Fig 2: Real-time sign language translation Block Diagram

Block Diagram Explanation

i). Camera Input
ii). Image Processing
iii). Gesture Recognition Model
iv). Text Generation
v). Text-to-Speech Conversion
vi). Audio Output

5. Modules Overview

i). Camera Input and Image Preprocessing Module
ii). Gesture Recognition Module
iii). Text and Speech conversion Module
iv). Output Display Module

Camera Input and Image Preprocessing Module

Description:

This module captures real-time video using a webcam or mobile camera. The video frames are continuously fed into the system for further processing.

Function

i). Live video capture
ii). Frame extraction
iii). Continuous input stream

Preprocessing improves image quality and removes noise to ensure accurate gesture recognition.

Steps

- Frame resizing
- Grayscale conversion
- Background removal
- Hand segmentation
- Normalization

Gesture Recognition Module

Description:

This is the core module that uses a CNN (Convolutional Neural Network) or MediaPipe + ML model to classify hand gestures.

Function:

- Feature extraction
- Gesture classification
- Mapping gestures to alphabets/words

Text and Speech Conversion Module

Description:

The recognized gesture is converted into readable text and displayed on the screen.

Function:

- Convert gesture labels to text
- Sentence formation
- Display text output

This module converts text into human-like speech using Text-to-Speech (TTS) engines such as Google TTS or Pyttsx3.

Function:

i). Text input
ii). Voice synthesis
iii). Audio generation

Output Display Module
Description:
Displays translated text and plays audio output for users.

Output:
- Text on screen
- Audio through speakers

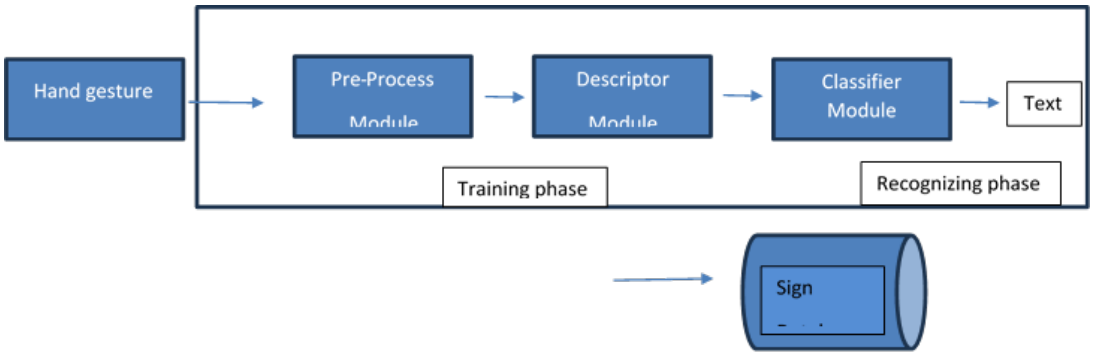

Fig 3: Static Gesture Recognition Core

Workflow Steps
i). User performs hand gesture
ii). Camera captures video
iii). Frames are preprocessed
iv). Gesture is recognized by ML model
v). Text is generated
vi). Audio output is played

Fig 4: Screenshots

Screens
- Camera capture screen
- Gesture detection screen
- Text output display
- Audio playing notification

6. Results and Discussion

The proposed system demonstrates accurate recognition of static hand gestures under controlled lighting conditions. The use of deep learning models improves recognition accuracy and response time. The system successfully converts recognized gestures into meaningful text and audio output, enabling effective real-time communication. Performance can be further improved by increasing training data diversity and incorporating dynamic gesture recognition.

7. Acknowledgment

The authors would like to express their sincere gratitude to Dr. K. Karuppasamy, Head of the Department of Computer Science and Engineering, RVS College of Engineering and Technology, for his continuous guidance, encouragement, and valuable suggestions throughout the development of this project. His support and motivation played a crucial role in the successful completion of the work.

We would like to thank all the faculty members of the Department of Computer Science and Engineering for their cooperation and support. We are also grateful to the management of RVS College of Engineering and Technology for providing the necessary facilities and resources to carry out this project successfully.

Finally, we would like to express our sincere appreciation to our friends and family members for their encouragement and moral support, which helped us stay motivated throughout the project duration.

8. Conclusion

The Real-Time Sign Language Translator to Text and Audio is an AI-based solution that helps bridge the communication gap between hearing-impaired individuals and society. It uses computer vision and machine learning to recognize sign language gestures in real time and converts them into accurate text output and audio speech using text-to-speech technology. This system promotes inclusivity, accessibility, and independence for deaf and mute individuals, and with further improvements in training data and accuracy, it has strong potential for real-world deployment.

References

1. Gonzalez RC, Woods RE. *Digital Image Processing*. 4th ed. Pearson; 2018.
2. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. MIT Press; 2016.
3. OpenCV. *Open Source Computer Vision Library*. https://opencv.org
4. MediaPipe. Google. https://mediapipe.dev
5. TensorFlow. https://www.tensorflow.org/