INTERNATIONAL JOURNAL OF ADVANCED INNOVATIVE TECHNOLOGY IN ENGINEERING

Published by Instigative Research Private Limited Home page: www.ijaite.co.in

Gesture Speak: An Intelligent Sign Language Recognition System

[1]Prof. Mrunali Gajbhiye, [2]Ms. Akanksha Dadwe, [3]Ms. Urvashi Selokar, [4]Ms. Arya Gawande, [5]Ms. Khushi Jaiswal

[1,2,3,4,5]Department of Computer Engineering, SCET, Nagpur, Maharashtra, India.

[1]mrunaligedam40@gmail.com, [2]uravashiselokar16@gmail.com
[3]aryagawande18@gmail.com, [4]khushdj001@gmail.com,
[5]akankshadadwe2005@gmail.com

Article History

Received on: 20 July 2026
Revised on15 Aug 2026
Accepted on: 20 Sept 2026
Keywords: Sign Language, Artificial Intelligence, Computer Vision, Machine Learning,, Media Pipe, OpenCV, TensorFlow, Accessibility, Real-Time Translation.

e-ISSN: 2455-6491

DOI: 10.65809/IJAITE/26/v03sp02/060

Production and hosted by
www.insightiveinc.org

©2026|All right reserved

ABSTRACT

This project aims to minimize communication difficulties between deaf or hearing-impaired individuals and people unfamiliar with sign language by introducing an AI-powered real-time translation system. The proposed solution captures hand movements through a webcam and analyzes them using computer vision and machine learning techniques. Tools such as OpenCV handle live video acquisition, Media Pipe detects hand landmarks, and TensorFlow/Keras classifies gestures into meaningful outputs. The extracted landmark features are used to train deep learning models like CNN or LSTM for reliable gesture recognition. A well-diversified dataset collected under varying lighting and background conditions improves the robustness and Communication plays a vital role in everyday human interaction, yet individuals adaptability of the model such as education, healthcare, and reducing accessibility and inclusivity. To address this issue, this paper presents an AI-based Sign Language Translator that recognizes hand gestures in real time and converts them into readable text or speech.

1. INTRODUCTION

Who are deaf or hearing-impaired often face challenges because sign language is not widely understood by the general population. This gap limits effective communication in education, workplaces, healthcare, and public environments, The proposed system integrates computer vision and machine learning techniques, using Open CV for video capture, Media Pipe for hand landmark detection, and Tensor Flow/Keras models such as CNN and LSTM for gesture classification. Extracted hand features are analyzed to identify gestures, which are then translated into meaningful outputs through a structured pipeline consisting of video capture, hand detection, feature extraction, classification, and text or speech generation. By enabling real-time gesture recognition, the system aims to reduce communication barriers, enhance accessibility, and promote inclusive interaction between sign language users and individuals unfamiliar with sign language, while ensuring accuracy, efficiency, and practical usability.

2. LITERATURE REVIEW

The literature survey highlights several significant contributions toward the development of sign language translation and gesture recognition systems. The work by Mohammed Fazal and Adharsh Reghu (2024) focused on a real-time sign language translator capable of converting recognized gestures into text or speech, enabling natural communication between hearing-impaired individuals and others; however, the system faced challenges in maintaining real-time performance due to processing delays.

Malay Kumar and S. Sarvajit Visagan (2024) proposed an enhanced translation framework connecting American Sign Language (ASL) and Indian Sign Language (ISL) using large language models, offering cross-language translation capability, though it required large and diverse datasets for effective training. K. Sharma and A. Kaushik (2022) provided a comprehensive analysis of technological approaches in sign language recognition, emphasizing their role in bridging communication gaps, but highlighted limitations such as reduced accuracy under varying lighting and background conditions. Y. S. Tan, K. M. Lin, and C. P. Lee (2021) introduced an enhanced densely connected convolutional neural network for hand gesture recognition, capable of identifying both simple and complex gestures, yet lacking the ability to interpret contextual or sentence-level meaning. Additionally, S. Stoll and N. C. Camgoz (2020) explored sign language production using neural machine translation to generate sign language videos from text or speech automatically, although the approach required high computational resources and long training time due to GPU-intensive processing. Overall, these studies demonstrate progress in gesture recognition and translation while highlighting challenges such as dataset requirements, computational complexity, real-time performance, and contextual understanding, motivating the need for efficient and practical real-time sign language translation systems.

Table 1: Literature Survey

Author(s)	Title	Merits	Demerits
Mohanmed Fazal & Adharsh Reghu (2024)	Real time sign language translator	Converts recognized signs into written or spoken language, allowing clear and natural communication with hearing users.	Real-time performance could suffer due to delays in video processing or model inference
Malay Kumar & S.Sarvaji t Visagan(2024)	Enhanced sign language translation between (ASL)&(ISL)using LLMs	Introduce s a unique framework that connects two distinct sign language feature not common in existing systems.	LLMs and gesture models need large, diverse datasets of ASL and ISL gestures to perform effectively
K Sharma & A Kaushik(2022)	A comprehensive analysis on technological approaches in sign language recognition	Bridges the communication gap between hearing/speech-impaired individuals and normal speakers, fostering social inclusion	Accuracy can drop if the lighting or background condition vary during gesture capture
YS Tan, KM Lin & CP lee(2021)	Hand gesture recognition via enhanced densely connected convolution neural network	Capable of recognizing both simple and complex hand gestures, suitable for sign language interpretation	Recognizes gestures but does not interpret context or sentence-level meaning sign language
S. Stoll and N. C. Camgoz (2020)	towards sign language production using neural machine translation	The system can automatically generate sign language videos directly from spoken or written sentences without human intervention	GAN-based video generation and motion Synthesis demand high GPU Power and long training time.

3. METHODOLOGY

Workflow starts with video acquisition, where a webcam captures live video using the OpenCV library, and the incoming frames are continuously analyzed to identify hand movements. Subsequently, hand detection is carried out with Media Pipe, which detects 21 hand landmarks corresponding to finger joints and palm orientation. These landmarks assist in accurately identifying the hand region and monitoring gesture motion.

Following hand detection, feature extraction is performed by gathering essential parameters such as finger coordinates, joint angles, hand values are converted into a numerical feature vector representing each gesture. The generated feature vectors are then used to build a dataset containing multiple samples for various gestures under different lighting conditions and

hand positions, improving the system's robustness and generalization capability.

After dataset preparation, the model training phase is conducted. Machine learning techniques such as Convolutional Neural Network (CNN) or Long Short-Term Memory (LSTM) are trained using the extracted features. During this stage, the model learns gesture patterns, updates weights through back propagation, and reduces prediction error to enhance accuracy. Once trained, the model is utilized for real-time gesture prediction. When a user performs a gesture, the webcam captures the frame, Media Pipe extracts landmarks, and the trained model predicts the corresponding gesture. Finally, the identified gesture is converted into text and optionally into speech output. The recognized text is displayed on the screen, and for audio output, Google Text-to-Speech (gTTS) is employed to produce speech. This end-to-end pipeline supports real-time communication by translating sign language gestures into understandable text or voice, thereby minimizing the communication gap between sign language users and non-signers. RELE based gesture recognition, where static hand images were used to identify alphabets and simple words. Although these approaches established the foundation for gesture-based communication, they were limited in handling dynamic gestures and lacked real-time capabilities. With the evolution of deep learning, Convolutional Neural Networks (CNNs) became popular for recognizing more complex gestures, improving overall accuracy and robustness; however, these models often required large datasets and were sensitive to variations in lighting and background conditions.

Subsequent research focused on real-time translation systems that utilized continuous video input along with advanced hand tracking mechanisms. The adoption of MediaPipe and landmark-based detection techniques enhanced gesture localization by identifying finger joints and palm orientation, enabling efficient feature extraction compared to traditional methods. Some studies also incorporated sequential models such as Long Short-Term Memory (LSTM) networks to capture temporal relationships in gesture sequences, allowing recognition of dynamic movements and short phrases. Despite these advancements, many systems demanded high computational resources and extensive training data for optimal performance.

More recent developments have emphasized multimodal frameworks that integrate gesture recognition with text-to-speech conversion to improve communication. These systems translate recognized gestures into both text and audio outputs, facilitating smoother interaction between hearing-impaired users and non-signers. Learning-based gesture classification, and real-time text and speech output to deliver a practical and accessible sign language translation system.

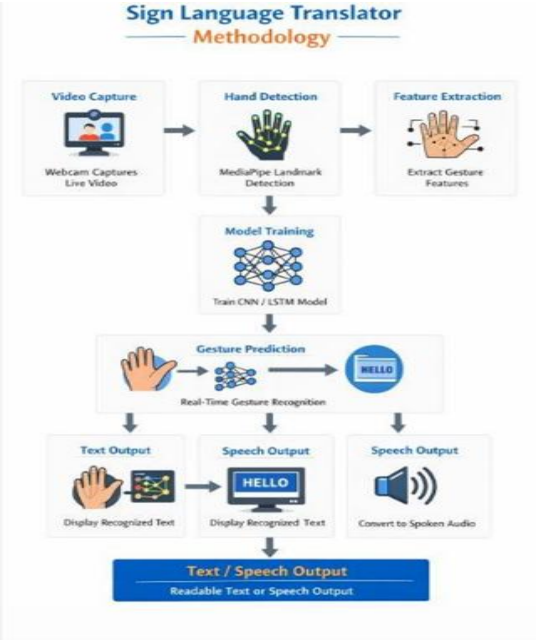

Figure 1: Sign Language Translator

4. IMPLEMENTATION

The implementation of the proposed AI-based Sign Language Translator is developed using Python by integrating computer vision, hand tracking, and machine learning techniques to recognize gestures in real time. The system is organized into several modules including video capture, hand detection, feature extraction, model training, gesture prediction, and text or speech output. Initially, the environment is prepared by installing essential libraries such as OpenCV,Media Pipe, NumPy, TensorFlow/Keras, and gTTS. These tools support video processing, hand landmark identification, data manipulation, model development, and speech synthesis.

The process starts with real-time video acquisition using OpenCV, where the webcam continuously captures frames that are forwarded to the Media Pipe hand tracking module. Media Pipe detects the hand region and extracts 21 landmark points corresponding to finger joints and palm orientation. These landmark coordinates are converted into feature vectors and normalized to maintain consistency across different hand positions and camera distances. The extracted features are then used to create a dataset for multiple gestures such as alphabets and commonly used words. Several samples are collected for each gesture under varying lighting conditions to enhance model accuracy.

After building the dataset, the machine learning model is implemented using TensorFlow/Keras. The dataset is divided into training and testing

subsets, and models such as Convolutional Neural Network (CNN) or Long Short-Term Memory (LSTM) are trained to learn gesture patterns. During training, the model updates its weights through back propagation and reduces classification error to improve performance. Once training is complete, the trained model is saved and later loaded for real-time inference. During prediction, live frames are captured, hand landmarks are extracted, transformed into feature vectors, and provided to the trained model to recognize the gesture.

Finally, the predicted gesture is displayed as text on the screen using OpenCV. For speech output, the recognized text is sent to the Google Text-to-Speech (gTTS) module, which converts the text into audio and plays it through system speakers. This complete implementation supports real-time gesture recognition and translation, enabling smooth communication between sign language users and individuals who are unfamiliar with sign language.

5. RESULT

The proposed AI-based Sign Language Translator was effectively implemented and evaluated using real-time webcam input. The system successfully detected hand gestures and converted them into readable text as well as speech output. Media Pipe accurately identified 21 hand landmarks, which supported precise feature extraction and enhanced gesture recognition performance. The trained machine learning model showed dependable classification of predefined gestures such as alphabets and commonly used words under varying lighting conditions and different hand orientations. During testing, the system continuously processed live video frames and generated predictions with very low latency, enabling smooth real-time interaction. The results indicated that the integration of OpenCV, Media Pipe, and TensorFlow/Keras enabled efficient gesture detection and classification. The predicted gestures were displayed on the screen in text form, and the gTTS module successfully produced the corresponding audio output. The system performed effectively in controlled indoor environments and maintained satisfactory accuracy when moderate changes in background and lighting conditions were present. Overall, the implementation achieved real-time gesture recognition, minimized communication barriers, and demonstrated the practicality of AI-based solutions for supporting communication between sign language users and individuals unfamiliar with sign language.

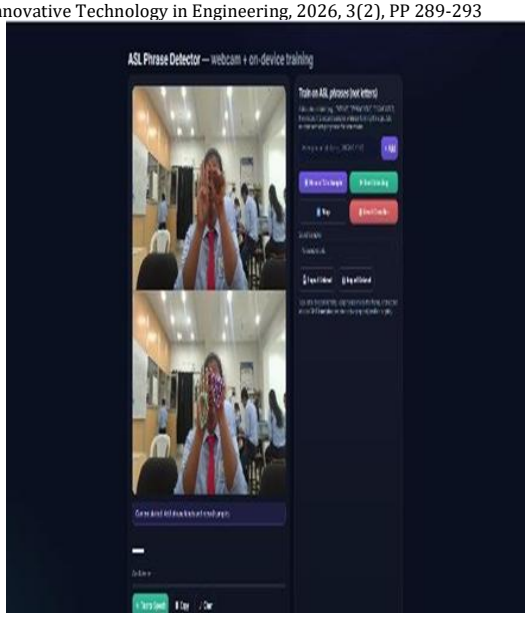

Figure 2: Real-Time Gesture Capture

CONCLUSION

The proposed AI-based Sign Language Translator presents an efficient approach for minimizing the communication barrier between deaf or hearing-impaired individuals and non-signers. By combining computer vision, hand tracking, and machine learning methods, the system can Recognize hand gestures in real time and translate them into both text and speech outputs. The integration of OpenCV for video acquisition, Media Pipe for hand landmark detection, and TensorFlow/Keras models for gesture classification contributed to accurate and reliable performance. The organized workflow including video capture, feature extraction, model training, and gesture prediction enabled consistent and smooth recognition of gestures. The developed system produced satisfactory accuracy across different lighting conditions and varying hand orientations while preserving real-time responsiveness with minimal delay. Providing both text and speech outputs improved usability and made the system applicable in practical communication situations. Overall, the proposed solution improves accessibility, encourages inclusive interaction, and highlights the effectiveness of AI-based technologies in supporting hearing-impaired individuals. Future enhancements may involve increasing the gesture dataset, enabling sentence-level gesture interpretation, and optimizing the model for deployment on mobile or embedded platforms to enhance portability and real-world usage.

REFERENCES

[1] ZainEldin , H. Gamel, S.A. Talaat, F.M. Aljohani, M. Baghdadi, N.A.; Malki, A.; Badawy, M. Elhosseini M.A. Silent no more: A comprehensive review of artificial intelligence, deep learning, and machine learning in

facilitating Deaf and mute communication. Artif. Intell. Rev, 2024.

[2] Alaftekin, M.Pacal, I. Cicek, K. Real-time sign language recognition based on YOLO algorithm. Neural Comput. Appl. 2024.

[3] Papatsimouli, M.Sarigiannidis, P.Fragulis, G.F. A survey of advancements in real-time sign language translators: Integration with IoT technology. Technologies 2023.

[4] Buttar, A.M.; Ahmad, U. Gumaei, A.H. Assiri, A. Akbar, M.A. Alkhamees, B.F. Deep learning in sign language recognition: A hybrid approach for the recognition of static and dynamic signs. Mathematics 2023

[5] A.K. Sahoo and R. Dash, 'A review on vision-based hand gesture recognition for human-computer interaction,' Journal of Ambient Intelligence and Humanized Computing, vol. 13, pp. 573–589, 2022.

[6] P. Dubey, "Sign language conversion flex sensor based on iot" International Journal of Research in Engineering and Science (IJRES), vol. 9, no. 2, pp. 69–71, 2021.

[7] P. K. Sahu and A. Tripathi, 'AI-powered communication tools for the deaf community,' IEEE Access, vol. 9, pp. 145632–145645, 2021.

[8] Wadhawan, A., Kumar, P. Deep learning-based sign language recognition system for static signs. Neural Comput & Applic 32, 7957–7968 (2020). https://doi.org/10.1007/s00521-019-04691-y