

# GestureMate: An AI-Driven System for Real-Time Malayalam Sign Language and Speech Translation

Jeswin Sabu

*Computer Science and Engineering*  
St. Joseph's College of Engineering and Technology  
Palai, India  
jeswinsabu2026@cs.sjcetpalai.ac.in

Kevin Biju Kulangara

*Computer Science and Engineering*  
St. Joseph's College of Engineering and Technology  
Palai, India  
kevinbijukulangara2026@cs.sjcetpalai.ac.in

Prapanch J

*Computer Science and Engineering*  
St. Joseph's College of Engineering and Technology  
Palai, India  
prapanchj2026@cs.sjcetpalai.ac.in

Stephin Mathew

*Computer Science and Engineering*  
St. Joseph's College of Engineering and Technology  
Palai, India  
stephinmathew2026@cs.sjcetpalai.ac.in

Vimal Babu P

*Assistant Professor, Department of CSE*  
St. Joseph's College of Engineering and Technology  
Palai, India  
vimalbabu@sjcetpalai.ac.in

**Abstract**—Communication barriers between deaf-mute individuals and hearing people continue to limit inclusive interaction in everyday environments such as healthcare facilities, workplaces, and educational institutions. While sign language serves as a primary means of expression for deaf-mute users, the lack of widespread sign language knowledge among hearing individuals creates significant challenges. To address this issue, this paper presents GestureMate, an AI-driven assistive system designed for real-time translation of Malayalam Sign Language into text and speech, while also converting spoken responses into readable text. The proposed system integrates computer vision-based hand gesture recognition, speech-to-text processing, and text-to-speech synthesis to enable seamless bidirectional communication. GestureMate captures hand gestures using a camera, extracts key hand landmarks, and classifies gestures into meaningful text representations. These text outputs are further converted into audible speech for hearing users. Conversely, spoken input from hearing individuals is transcribed into text, allowing deaf-mute users to understand verbal responses instantly. A distinctive feature of the system is its custom sign training module, which allows users to define personalized or regional sign variations, making the system adaptable to diverse communication needs. Additionally, GestureMate includes a sign language learning module to promote awareness and accessibility among hearing users. Experimental evaluation demonstrates that the system provides accurate gesture recognition and low-latency speech transcription, enabling smooth real-time interaction.

**Index Terms**—Malayalam Sign Language, Gesture Recognition, Speech-to-Text, Text-to-Speech, Assistive Communication Systems, Artificial Intelligence.

## I. INTRODUCTION

Communication is a fundamental aspect of human interaction; however, individuals who are deaf-mute often face significant challenges while communicating with hearing people

in everyday situations. Although sign language serves as an effective medium of expression for the deaf-mute community, the lack of sign language knowledge among the general population creates a major communication gap. This issue becomes more critical in environments such as hospitals, educational institutions, workplaces, and public service centers, where timely and accurate communication is essential.

Sign language translation systems have been actively researched to address this gap by converting hand gestures into textual or spoken output. Moreover, many systems lack adaptability to regional sign variations, making them less effective for localized communication needs. In India there are various regional sign languages such as Malayalam Sign Language, yet they remain underrepresented in intelligent translation systems.

Recent advancements in artificial intelligence, particularly in computer vision and speech processing, have enabled the development of real-time assistive communication systems. Vision-based hand gesture recognition techniques allow systems to identify hand landmarks and movements using standard cameras, eliminating the need for additional hardware. Similarly, modern speech-to-text and text-to-speech technologies enable accurate transcription and natural voice synthesis, facilitating seamless bidirectional communication. Despite these advancements, integrating gesture recognition and speech processing into a unified, real-time system that supports regional sign languages remains a challenging task.

To address these limitations, this paper presents GestureMate, an AI-driven assistive communication system designed for real-time translation of Malayalam Sign Language into

text and speech, while also converting spoken input from hearing users into readable text. The proposed system enables natural two-way interaction between deaf-mute and hearing individuals using a multimodal approach. *GestureMate* employs computer vision techniques for hand gesture detection and classification, speech recognition for transcribing spoken responses, and text-to-speech synthesis for generating audible output. A key feature of the system is its custom sign training capability, which allows users to define personalized or regional sign variations, improving adaptability and inclusivity.

The primary contributions of this work are as follows: (i) the design of a real-time multimodal communication framework tailored for Malayalam Sign Language, (ii) the integration of gesture recognition, speech-to-text, and text-to-speech technologies into a single system, and (iii) the introduction of a customizable sign training module to support regional and user-specific sign variations.

The remainder of this paper is organized as follows. Section II reviews existing work related to sign language recognition and speech-based communication systems. Section III describes the architecture and components of the proposed system. Section IV explains the methodology used for gesture recognition and speech processing. Section V presents the experimental results and discussion. Finally, Section VI concludes the paper and outlines future research directions.

## II. RELATED WORK

Research on sign language recognition has gained considerable attention over the past decade due to its potential to improve communication accessibility for deaf-mute individuals. Early sign language translation systems primarily relied on sensor-based devices such as data gloves and motion sensors to capture hand movements accurately [1]. Although these approaches provided precise gesture information, their dependency on specialized hardware limited portability and real-world applicability.

With advancements in computer vision, vision-based sign language recognition systems emerged as a practical alternative. These systems use cameras to capture hand gestures and apply image processing and machine learning techniques to extract meaningful features [2]. Recent studies have demonstrated that convolutional neural networks and landmark-based hand tracking methods can achieve effective real-time gesture recognition without requiring additional hardware [3]. However, most of these systems are trained on widely used sign languages and often fail to address regional or language-specific sign variations.

Several researchers have explored the integration of sign language recognition with speech processing technologies to enable bidirectional communication. Speech-to-text systems have been extensively studied for transcribing spoken language into textual form using deep learning models [4]. Similarly, text-to-speech synthesis has evolved with neural speech models capable of generating natural and intelligible voice output [5]. Despite these advancements, existing systems typically

focus on isolated components and lack seamless integration for real-time assistive communication.

In the Indian context, limited research has focused on regional sign languages, and very few systems support Malayalam Sign Language. Existing solutions often lack adaptability to personalized or localized sign variations, which reduces their effectiveness in practical scenarios [6]. Additionally, most systems do not provide customization or learning features that could enhance usability for both deaf-mute and hearing users.

In contrast, the proposed *GestureMate* system aims to address these limitations by providing a real-time multimodal communication framework specifically designed for Malayalam Sign Language. By integrating vision-based gesture recognition, speech-to-text transcription, and text-to-speech synthesis within a unified platform, *GestureMate* enables seamless two-way communication. Furthermore, the inclusion of a custom sign training module allows users to define regional and personalized gestures, enhancing adaptability and inclusivity beyond existing approaches.

## III. PROPOSED SYSTEM

The proposed system, named *GestureMate*, is an AI-driven assistive communication platform designed to enable real-time bidirectional interaction between deaf-mute and hearing individuals using Malayalam Sign Language. The system integrates multiple intelligent modules that work together to translate hand gestures into text and speech, while also converting spoken input into readable text. The overall design emphasizes real-time performance, accessibility, and adaptability to regional and personalized sign variations.

*GestureMate* operates using a multimodal communication approach. For deaf-mute users, the system captures hand gestures through a standard camera and processes them using computer vision techniques to identify meaningful signs. These recognized signs are converted into textual representations, which can then be synthesized into audible speech for hearing users. Conversely, spoken input from hearing users is captured through a microphone and transcribed into text, allowing deaf-mute users to understand verbal responses instantly. This bidirectional flow ensures seamless interaction without requiring either user to have prior knowledge of sign language.

The architecture of *GestureMate* consists of several interconnected modules. The Sign Language Detection Module is responsible for capturing real-time video input and identifying hand landmarks and movements associated with Malayalam sign gestures. The extracted features are classified into corresponding textual representations using trained recognition models. The Speech Recognition Module processes spoken language input and converts it into text for display on the user interface. To facilitate communication from deaf-mute users to hearing users, the Text-to-Speech Conversion Module transforms recognized text into natural-sounding voice output.

A unique aspect of the proposed system is the Custom Sign Training Module. This module allows users to record and assign meanings to new gestures, enabling the system to

learn personalized or region-specific sign variations. By incorporating user-defined gestures into the recognition pipeline, GestureMate improves adaptability and supports diverse communication styles. Additionally, the Sign Language Learning Module provides tutorials and demonstrations to help hearing users learn basic sign language, thereby promoting inclusivity and awareness.

The User Interface Module serves as the central interaction layer, providing an intuitive and accessible environment for both deaf-mute and hearing users. It enables users to switch between communication modes, view real-time text transcriptions, listen to audio outputs, and access learning or training features. The interface is designed to ensure ease of use, minimal latency, and clear visual feedback during communication.

Overall, the proposed GestureMate system combines gesture recognition, speech processing, and user-centered design to create an effective and inclusive communication solution. By focusing on Malayalam Sign Language and supporting customization through user-defined gestures, the system addresses key limitations of existing approaches and offers a practical solution for real-world assistive communication.

#### IV. METHODOLOGY

This section describes the methodology adopted for implementing the GestureMate system. The proposed approach integrates computer vision techniques for gesture recognition, speech processing for verbal input, and text-to-speech synthesis for audio output. The overall methodology is designed to support real-time performance while ensuring accuracy and adaptability for Malayalam Sign Language translation.

##### A. Gesture Recognition

Hand gesture recognition is performed using a vision-based approach that combines OpenCV and the MediaPipe framework. A standard camera is used to capture real-time video input of hand gestures performed by the deaf-mute user. OpenCV is employed for video acquisition, frame preprocessing, and image normalization [8]. Each frame is resized and processed to reduce noise and improve detection stability under varying lighting conditions.

MediaPipe Hands is utilized to detect and track hand landmarks in real time. The framework identifies key points of the hand, including finger joints and palm coordinates, which serve as feature representations for gesture classification [7]. These landmarks are extracted from each frame and structured into feature vectors. The extracted features are then analyzed to classify the performed gesture into a corresponding Malayalam sign. Once recognized, the gesture is mapped to its predefined textual representation. This vision-based approach eliminates the need for specialized hardware and enables real-time gesture detection using commonly available devices.

##### B. Speech-to-Text

To enable communication from hearing users to deaf-mute users, the system incorporates Windows Speech-to-Text technology. Spoken input is captured through a microphone and

processed using the built-in speech recognition capabilities provided by the Windows operating system. The speech recognition engine converts verbal input into textual form, which is then displayed on the user interface.

Windows Speech Recognition is selected due to its reliability, ease of integration, and low-latency performance in desktop environments. The module supports continuous speech recognition and provides accurate transcription for conversational input. By leveraging native operating system services, the system avoids additional computational overhead and ensures seamless real-time speech transcription for hearing users [9].

##### C. Text-to-Speech

The Text-to-Speech (TTS) component of GestureMate is implemented using Cloud Text-to-Speech technology. Once a gesture is recognized and converted into text, the textual output is passed to the Cloud TTS module to generate the corresponding audio output. This enables hearing users to understand the communication from deaf-mute users in spoken form.

Cloud TTS produces natural-sounding speech with clear pronunciation and smooth intonation. The generated audio output is played through the system's audio interface with minimal delay, ensuring a smooth conversational experience. The use of Cloud TTS allows the system to deliver expressive and intelligible voice responses, improving the overall quality of interaction between users [10].

##### D. System Integration and Workflow

The complete workflow of the system begins with gesture or speech input, depending on the user type. Gesture input is processed through the gesture recognition pipeline and converted into text, which is then synthesized into speech using Cloud TTS. Conversely, spoken input from hearing users is transcribed into text using Windows Speech-to-Text and displayed to the deaf-mute user. These modules are integrated through a centralized control flow that ensures synchronized processing and real-time feedback.

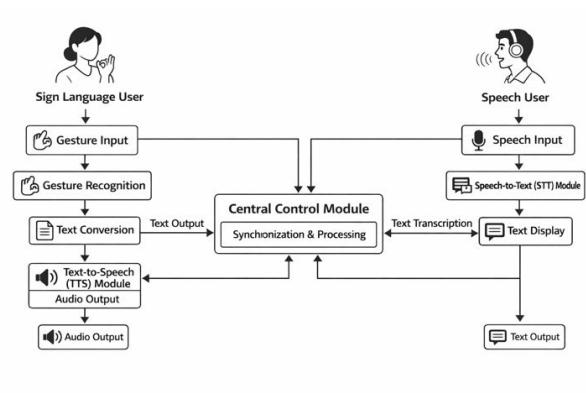


Fig. 1. System workflow of the proposed GestureMate communication framework

As shown in Fig. 1, the system follows a modular architecture where gesture recognition, speech processing, and output generation operate independently while contributing to a unified communication pipeline. This design ensures efficient data flow, low latency, and adaptability to real-world communication scenarios.

## V. RESULTS AND DISCUSSION

This section presents the experimental results and performance analysis of the proposed *GestureMate* system. The evaluation focuses on gesture recognition accuracy, speech transcription performance, response latency, and overall system usability during real-time interaction between deaf-mute and hearing users.

The gesture recognition module was tested using a set of commonly used Malayalam sign gestures performed under varying lighting conditions and hand orientations. The OpenCV and MediaPipe-based approach demonstrated reliable hand landmark detection and stable gesture recognition in real-time scenarios. The system achieved consistent recognition performance for predefined gestures, with improved accuracy observed when gestures were performed within the camera's optimal field of view.

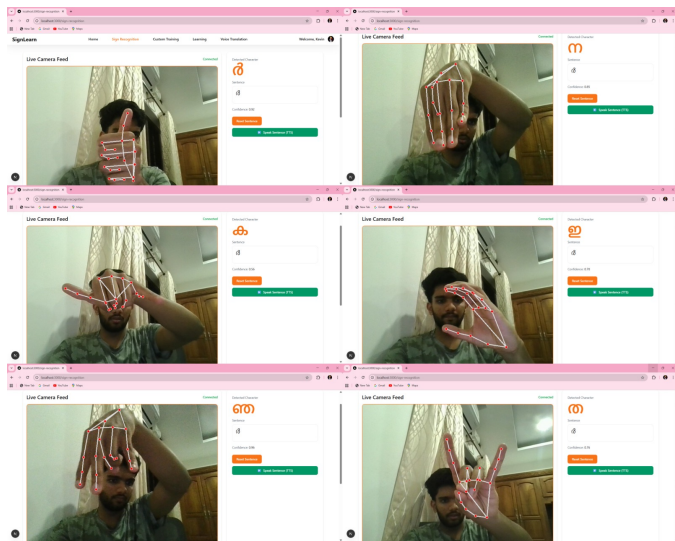


Fig. 2. Sample gesture recognition outputs with detected hand landmarks and classified Malayalam characters

As shown in Fig. 2, the system accurately detects hand landmarks and classifies different Malayalam sign gestures in real time. The visual results highlight the robustness of the MediaPipe-based hand tracking framework under varying hand orientations and background conditions. Minor variations in hand positioning and lighting had limited impact on recognition performance, indicating the stability of the proposed gesture recognition approach.

The speech-to-text module using Windows Speech Recognition was evaluated with conversational speech input from hearing users. The module provided accurate and near-instantaneous transcription for clear speech, enabling smooth

communication flow. Transcription accuracy was observed to be higher in quiet environments, while minor delays or misinterpretations occurred in the presence of background noise. Nevertheless, the performance was sufficient for real-time conversational use.

The text-to-speech module implemented using Cloud TTS produced natural and intelligible voice output with minimal delay. The generated speech output effectively conveyed the recognized text to hearing users, ensuring clarity during communication. The low latency of the TTS module contributed to a seamless interaction experience and prevented noticeable communication gaps.

End-to-end system latency was measured from gesture or speech input to final output display or audio playback. The system demonstrated low response time, enabling real-time bidirectional communication. The modular architecture allowed parallel processing of gesture recognition and speech synthesis, further reducing overall latency.

Despite its effectiveness, the system has certain limitations. Gesture recognition accuracy may decrease when multiple hands appear simultaneously in the camera frame or when gestures are performed rapidly. Similarly, speech recognition performance depends on environmental noise levels and microphone quality. These limitations indicate potential areas for improvement through advanced gesture classification models and noise-robust speech processing techniques.

Overall, the experimental results demonstrate that *GestureMate* effectively enables real-time Malayalam sign language translation and speech-based communication. The system shows strong practical usability in real-world scenarios and provides a solid foundation for future enhancements in assistive communication technologies.

## VI. CONCLUSION AND FUTURE WORK

This paper presented *GestureMate*, an AI-driven assistive communication system designed to enable real-time translation of Malayalam Sign Language and spoken communication between deaf-mute and hearing individuals. The proposed system integrates vision-based gesture recognition using OpenCV and MediaPipe, speech-to-text conversion using Windows Speech Recognition, and text-to-speech synthesis using Cloud TTS. By combining these technologies into a unified framework, *GestureMate* enables seamless bidirectional communication without requiring specialized hardware.

The experimental results demonstrate that the system achieves reliable gesture recognition, accurate speech transcription, and natural-sounding speech output with low latency, making it suitable for real-time conversational use. A key strength of the system is its modular and adaptable design, which supports personalized gesture training and promotes inclusivity through sign language learning features. The focus on Malayalam Sign Language addresses an important gap in existing assistive communication systems and highlights the relevance of region-specific solutions.

Future work will focus on enhancing gesture recognition accuracy by incorporating advanced deep learning models

capable of handling complex and dynamic gestures. The system can be extended to support additional regional sign languages and multilingual speech processing to broaden its applicability. Further improvements may include noise-robust speech recognition, mobile platform deployment, and cloud-based scalability to support a larger user base. These enhancements will strengthen the system's potential as a practical and inclusive communication solution for diverse real-world environments.

#### REFERENCES

- [1] T. Starner and A. Pentland, "Real-time American Sign Language recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371–1375, Dec. 1998.
- [2] S. Mitra and T. Acharya, "Gesture recognition: A survey," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 3, pp. 311–324, May 2007.
- [3] C. Zhang, F. Zhang, and L. Zhao, "Vision-based hand gesture recognition using deep learning," *International Journal of Computer Vision and Signal Processing*, vol. 8, no. 2, pp. 45–52, 2018.
- [4] A. Graves, A. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver, BC, Canada, 2013, pp. 6645–6649.
- [5] Y. Wang *et al.*, "Tacotron: Towards end-to-end speech synthesis," in *Proc. Interspeech*, Stockholm, Sweden, 2017, pp. 4006–4010.
- [6] S. K. Sahoo and M. R. Reddy, "An overview of Indian sign language recognition systems," *International Journal of Computer Applications*, vol. 975, no. 8887, pp. 1–6, 2016.
- [7] C. Lugaresi *et al.*, "MediaPipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
- [8] OpenCV Team, "OpenCV: Open Source Computer Vision Library," [Online]. Available: <https://opencv.org/>. Accessed: 2025.
- [9] Microsoft Corporation, "Windows Speech Recognition Documentation," [Online]. Available: <https://learn.microsoft.com/en-us/windows/apps/design/input/speech-recognition>. Accessed: 2025.
- [10] Google Cloud, "Cloud Text-to-Speech Overview," [Online]. Available: <https://cloud.google.com/text-to-speech/>. Accessed: 2025.