IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510126

AI Powered Sign Language to Text and Speech

Samanvi I[1], Sanjana Hegde[1], Shravya K V[1], Shravya N Shetty[1], Bharathi K[2]

Students, Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, India[1]

Assistant Professor, Department of Computer Science and Engineering, Vivekananda College of Engineering and Technology, Puttur, India[2]

Abstract: Communication between individuals who use sign language and those unfamiliar with it can be challenging, particularly in everyday interactions. This paper presents an AI-powered sign language recognition system designed to convert hand gestures into text and speech in Kannada, Hindi, and Tamil. The proposed system uses a webcam to capture static hand gestures and MediaPipe to detect hand landmarks and extract spatial features. A Multilayer Perceptron (MLP) model classifies the extracted features into supported sign characters. The recognized characters are then combined to form words and sentences using language-specific text composition rules. Large Language Model (LLM) based text refinement is incorporated to improve the grammatical correctness and readability of the generated text. The refined text is displayed through a user interface and converted into speech using a Text-to-Speech (TTS) module. The proposed system demonstrates the potential of computer vision, machine learning, and natural language processing in supporting multilingual assistive communication and improving accessibility for sign language users.

Keywords: Artificial Intelligence, Sign Language Recognition, Gesture Recognition, Text to Speech, Large Language Model

I. INTRODUCTION

Communication is essential for everyday interaction. However, communication barriers can arise when sign language users interact with people who are unfamiliar with sign language. Technology-based sign recognition systems can help reduce these barriers by converting recognized gestures into text and speech.

This paper presents an AI-powered multilingual sign language detection system designed to recognize supported static hand gestures and convert them into text and speech in Kannada, Hindi, and Tamil. The system uses a webcam to capture hand gestures, MediaPipe to detect hand landmarks, and a Multilayer Perceptron (MLP) model to classify the extracted hand features.

The recognized characters are processed sequentially and combined into words and sentences using language-specific text composition rules. Large Language Model (LLM)-based text refinement is used to improve the readability and grammatical correctness of the generated text. A Text-to-Speech (TTS) module converts the refined text into spoken output. The system also provides a user interface for displaying recognized text and interacting with the recognition process.

Unlike systems designed to translate continuous sign-language movements, the proposed system focuses on static fingerspelling, in which individual characters are recognized and combined to construct words and sentences. This work explores the application of computer vision, machine learning, and natural language processing to support multilingual assistive communication.

II. RELATED WORK

Previous studies have explored different approaches to sign language recognition using computer vision and machine learning. MediaPipe-based methods have gained attention because they extract hand landmarks from webcam images, allowing gesture recognition using numerical features rather than processing every image directly. Varanasi *et al.* [1] investigated isolated Indian Sign Language recognition using MediaPipe landmarks, highlighting the use of landmark-based representations for gesture analysis.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510126

Machine learning classifiers have also been explored for recognizing static hand gestures. Researchers have combined MediaPipe landmark extraction with classification techniques to identify individual signs and convert them into text [2]. Such approaches are relevant to fingerspelling systems, where each hand gesture represents a character that can be recognized independently. Other research has investigated the combination of three-dimensional hand geometry and a Multilayer Perceptron (MLP) with MediaPipe-based hand skeleton detection for real-time American Sign Language recognition [3].

Several studies have extended gesture recognition systems to produce speech output. Vision-based gesture-to-speech approaches demonstrate how recognized gestures can be converted into text and subsequently spoken using text-to-speech technologies [4]. Recent research has also explored Large Language Model (LLM)-based natural language generation in sign language-to-speech systems, highlighting the potential of language models to improve the naturalness of generated communication [5].

Although these studies demonstrate progress in gesture recognition and assistive communication, challenges remain in multilingual character recognition, accurate character composition, text refinement, and integration of speech output. The proposed system addresses these aspects by combining MediaPipe-based hand-landmark extraction, MLP-based static character classification, language-specific text composition, LLM-based text refinement, and text-to-speech output for Kannada, Hindi, and Tamil. Unlike continuous sign language translation approaches, the proposed system focuses on recognizing individual static hand gestures and combining the recognized characters into words and sentences.

III. PROPOSED METHOD

The proposed system follows a sequence of processing stages to recognize static hand gestures and convert them into text and speech. The main stages are webcam input, hand landmark extraction, gesture classification, text composition, text refinement, and speech generation.

A. Image Acquisition

The webcam captures the user's hand gestures in real time. The captured frames are processed by the application and supplied to the hand detection component for further analysis.

B. Hand Landmark Detection and Feature Extraction

MediaPipe detects the hand and identifies its landmark points, including the wrist, finger joints, and fingertips. The landmark coordinates represent the spatial configuration of the hand and provide numerical features for gesture classification.

C. Gesture Classification

The extracted hand features are passed to a trained Multilayer Perceptron (MLP) model. The model predicts the corresponding character from the supported gesture classes. This process is repeated as the user presents successive hand gestures.

D. Text Composition and Refinement

The predicted characters are passed to the text composition component, which combines them sequentially into words and sentences according to language-specific rules. Text-processing and dictionary-based spelling-correction techniques are then applied to help identify and correct spelling errors in the generated text.

E. Speech Generation

The final text is displayed in the application interface and passed to the implemented Text-to-Speech (TTS) component. The TTS component generates spoken output from the recognized text, providing an additional communication channel.

 This work is licensed under a Creative Commons Attribution 4.0 International License

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510126

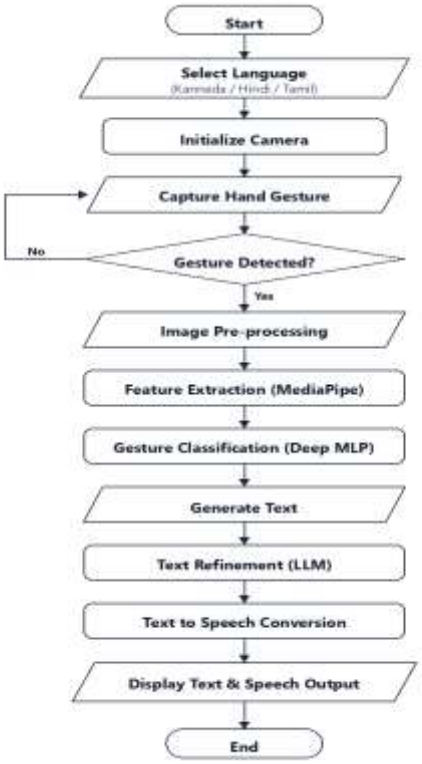


Fig. 1. System Architecture

IV. IMPLEMENTATION AND RESULTS

The proposed system converts static hand gestures into text and speech in Kannada, Hindi, and Tamil. A webcam captures the hand gestures, and MediaPipe detects hand landmarks. A trained Multilayer Perceptron (MLP) model uses the extracted features to recognize individual characters. The recognized characters are combined into words and sentences using language-specific text composition rules.

The generated text is displayed on the user interface and refined using Large Language Model (LLM)-based text processing to improve readability and grammatical correctness. The refined text is converted into spoken output using a Text-to-Speech (TTS) module. The interface also allows users to select a language, add recognized characters, and clear the generated text.

A. Testing and Evaluation

ID	Test Case	Expected Outcome	Actual Outcome	Result
1	Camera Initialization	Webcam initializes successfully	Webcam feed is displayed successfully	Pass
2	Real-Time Frame Capture	Live video frames are captured	Hand gestures are captured through the webcam	Pass
3	Hand Landmark Detection	MediaPipe detects hand landmarks	Hand landmarks are detected when the hand is visible	Pass
4	Gesture Classification	MLP model predicts the corresponding character	Predicted characters are displayed for supported gestures	Pass

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026
DOI: 10.17148/IJARCCE.2026.1510126

5	Character Composition	Characters are combined to form words and sentences	Characters are combined using language-specific rules	Pass
6	Text Refinement	Generated text is refined for readability and grammatical correctness	Refined text is displayed when processing is successful	Pass
7	Text-to-Speech	Generated text is converted into speech	Spoken output is produced for the selected language	Pass

B. Discussion and Limitations

The functional testing evaluates the major components of the proposed system, from gesture capture to speech generation. The system supports character-by-character composition, allowing individual recognized gestures to be combined into words and sentences. Language selection enables users to work with Kannada, Hindi, and Tamil.

However, recognition performance may vary depending on lighting conditions, hand visibility, hand positioning, and similarities between gestures. Incorrectly recognized characters may also affect the generated text and its refinement. Speech generation depends on the availability and correct operation of the selected language's speech engine. Further evaluation using documented test datasets, classification accuracy, and additional functional tests is required to establish the overall performance and reliability of the system

V. CONCLUSION

This work presents an AI-powered sign language recognition system that converts static hand gestures into text and speech in Kannada, Hindi, and Tamil. The system uses MediaPipe for hand landmark detection and a Multilayer Perceptron (MLP) model for character recognition. The recognized characters are combined into words and sentences using language-specific text composition rules. Large Language Model (LLM)-based text refinement is incorporated to improve the readability and grammatical correctness of the generated text, while a Text-to-Speech (TTS) module provides spoken output. The system integrates computer vision, machine learning, multilingual text processing, and speech synthesis to support accessible communication. The developed system provides a foundation for further improvements in multilingual static sign language fingerspelling and assistive communication.

Scope for Future Work

Future work includes:

- Improving character recognition accuracy under different lighting conditions and hand positions.
- Expanding the supported character sets and vocabulary in Kannada, Hindi, and Tamil.
- Enhancing language-specific text composition and LLM-based text refinement.
- Improving speech generation and pronunciation across all three languages.

ACKNOWLEDGMENT

We express our sincere gratitude to our project guide, Prof. Bharathi K, for her valuable guidance, encouragement, and support throughout this work. We thank Prof. Pradeep Kumar K G, Head of the Department of Computer Science and Engineering, and Dr. Purushothama Chippar, Principal, for their encouragement and for providing the necessary facilities. We also extend our thanks to the college management and all teaching and non-teaching staff members of the department for their support in completing this project

REFERENCES

[1] A. Bharadwaj Varanasi, M. Sinha, and T. Dasgupta, "Isolated Sign Language Recognition via MediaPipe Landmarks: A Case Study on Indian Sign Language," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2026, pp. 10595–10604

[2] "Real-Time Sign Language to Text Conversion Using MediaPipe Hand Landmarks and Random Forest Classifier," in 2026 6th International Conference on Trends in Material Science and Inventive Materials (ICTMIM), 2026, doi: 10.1109/ICTMIM68190.2026.11507737.

 This work is licensed under a Creative Commons Attribution 4.0 International License



International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 10, October 2026

DOI: 10.17148/IJARCCE.2026.1510126

[3] "3D Geometric Features Based Real-Time American Sign Language Recognition Using PointNet and MLP with MediaPipe Hand Skeleton Detection," Measurement: Sensors, vol. 38, 101697, 2025, doi: 10.1016/j.measen.2024.101697.

[4] "Vision-Based Real-Time Gesture-to-Speech Translation for Sign Language," Procedia Computer Science, vol. 258, pp. 2050–2059, 2025, doi: 10.1016/j.procs.2025.04.455.

[5] S. Suhartono et al., "The SmartGlove BISINDO Wearable Sensor and Large Language Model-Based Sign Language-to-Speech System for Users With Communication Disorders: Design Science Research Study," JMIR Rehabilitation and Assistive Technologies, vol. 13, 2026, e96258, doi: 10.2196/96258.

 This work is licensed under a Creative Commons Attribution 4.0 International License