



REAL-TIME SIGN LANGUAGE RECOGNITION SYSTEM

Bhong Sakshi Suresh¹

¹ Dattakala Group of Institutions, Faculty of Engineering, India

Corresponding author: sakshibhong0@gmail.com

Guide / Supervisor: Bere S. S.

ARTICLE ID IJREK-2026-00070	ARTICLE TYPE Research Article	RESEARCH AREA Computer Science	PUBLICATION Volume 1, Issue Issue 1 (2026)
RECEIVED 15 October 2026	REVISED 19 October 2026	ACCEPTED 18 October 2026	PUBLISHED 19 October 2026

Abstract

Communication between deaf or hard-of-hearing people and people unfamiliar with sign language can be challenging. This paper presents a proposed Real-Time Sign Language Recognition System that uses a camera to capture hand gestures and computer-vision techniques to classify signs and display corresponding text. The system is organized into video acquisition, hand-region processing, feature extraction, gesture classification, and textoutput modules. A practical implementation may use Python, OpenCV, MediaPipe for hand-landmark detection, and a trained machine-learning classifier. The interface is intended to provide immediate visual feedback and can be extended to speech output. Recognition performance depends on the sign vocabulary, lighting, camera quality, signing speed, and diversity of the training data. Evaluation should use a documented dataset and measured test results; no experimental results are claimed in this sample paper.

Keywords: Sign language recognition; real-time system; computer vision; hand gesture; MediaPipe; machine learning; accessibility. 2

I. INTRODUCTION

A. Background / Context Sign language is an important means of communication for many deaf and hard-of-hearing people. However, communication can be difficult when the other person does not understand the sign language being used. Computer vision and machine learning offer ways to recognize visible hand gestures and convert selected signs into readable text. A real-time recognition system processes camera frames continuously and provides feedback with limited delay. B. Problem Statement Many communication settings lack an interpreter or a shared sign language. Manual interpretation is not always immediately available, while text-based communication may interrupt natural interaction. A useful assistive system should detect hand gestures from a live camera stream, classify supported signs, and present results in a clear interface. Recognition can be affected by occlusion, background clutter, different users, and similar-looking signs. C. Objective / Aim The objective is to design a real-time sign language recognition system that captures video, detects hand landmarks or regions, extracts useful features, classifies supported signs, and displays the predicted label. The proposed modular design supports testing and future expansion of the vocabulary. D. Significance / Contribution The proposed system aims to support communication accessibility by converting a limited set of recognized signs into text. Its contribution is a clear, modular pipeline that can run on a standard computer with a camera and can be extended with additional signs, confidence feedback, or text-to-speech output. E. Paper Organization The remainder of this paper discusses related work, the proposed system and methodology, implementation choices, evaluation design, limitations, future scope, and conclusion.

II. LITERATURE REVIEW / RELATED WORK

A. Overview of Existing Research Research on sign language recognition includes sensor-based gloves, image-processing methods, hand-shape descriptors, deep-learning models, and landmark-based approaches. Camera-based systems are attractive because they do not require users to wear specialized hardware. Hand-landmark tools can represent fingers and joints as coordinates, which can be used as input to a classifier. B. Summary of Key Research Works Earlier approaches often

used handcrafted features and classifiers such as support vector machines or random forests. More recent approaches use convolutional neural networks for image-based recognition and temporal models for sequences of signs. Landmark-based pipelines can reduce input dimensionality and may be practical for real-time prototypes, although performance depends on the dataset and the complexity of the signs. C. Identified Research Gap A key challenge is building a system that is responsive while remaining reliable across different signers, lighting conditions, camera positions, and backgrounds. Static hand poses do not represent every sign; some signs depend on movement, facial expression, or body posture. A prototype must clearly state which signs it supports and avoid implying that a small vocabulary provides full sign-language translation. D. Contribution of the Proposed Work This work proposes a modular camera-based system that detects hand landmarks, derives features, classifies supported signs, and displays the predicted label. The focus is on real-time interaction, low-cost hardware, transparent evaluation, and an architecture that can later support dynamic gestures and a larger vocabulary.

III. PROPOSED SYSTEM AND METHODOLOGY

TABLE I

Module	Description
Video Input	Captures live frames from the webcam.
Hand Detection	Detects hand landmarks or the hand region.
Feature Extraction	Normalizes landmarks and creates a feature vector.
Sign Classifier	Predicts a label from the supported sign vocabulary.
Temporal Smoothing	Reduces flicker across consecutive predictions.
Text Output	Displays the recognized sign and optional confidence.
User Interface	Shows the camera feed, status, and prediction.

TABLE II

Step	Operation
1	Capture a frame from the camera.
2	Detect the hand and estimate landmarks.
3	Normalize landmark coordinates and extract features.
4	Send features to the trained sign classifier.
5	Check prediction confidence and smooth consecutive labels.
6	Display the recognized sign or a neutral fallback message.
7	Repeat until the user exits and release camera resources.

Fig. 2. Query classification and response workflow. G. Advantages of the Proposed System • Live Camera Input: captures gestures from a webcam. • Hand Landmark Detection: identifies hand joints and finger positions. • Real-Time Prediction: classifies supported gestures as frames arrive. • Accessible Output: presents recognized signs as readable text. • Modular Design: allows the detector, classifier, and interface to be updated separately. • Future Extension: can add dynamic signs, regional sign vocabularies, and optional speech output.

IV. IMPLEMENTATION DETAILS

TABLE III

Component	Specification / Description
Frontend	Python GUI or a lightweight web interface
Video Processing	OpenCV
Hand Landmarks	MediaPipe Hands or equivalent detector
Classifier	SVM, Random Forest, or a validated neural network
Dataset	Labeled examples for each supported sign
Deployment	Computer with webcam; hardware details to be documented

TABLE IV

Input	Processing	Output
Camera frame	Hand detection → landmark normalization	Hand landmarks/features
Static sign	Feature vector → trained classifier	Predicted sign label
Low-confidence frame	Confidence threshold / smoothing	Neutral or previous stable output
No hand detected	Detection check	Prompt or blank prediction

V. RESULTS AND ANALYSIS

TABLE V

Test ID	Input Type	Expected Behaviour
T1	Supported static sign	Display the correct trained label when recognized.
T2	No hand in frame	Show a neutral message or no prediction.
T3	Unseen/ambiguous gesture	Avoid presenting a low-confidence guess as certain.
T4	Changing lighting/background	Continue processing; document recognition errors.
T5	Camera unavailable	Show an error message and exit safely.

TABLE VI

Metric	Measured Value
Classification Accuracy	To be measured
Per-Class Precision / Recall	To be measured
Average Inference Latency	To be measured
Processing Frame Rate	To be measured

TABLE VII

Parameter	Manual Communication	Sensor/Glove-Based	Proposed Camera System
Hardware	No electronic device	Instrumented glove	Standard camera/webcam
Input Method	Person interprets signs; partner	Finger/bend sensors	Video and hand landmarks
Output	Depends on language	Mapped text/speech	Recognized label as text
Portability	No setup required	Wearable hardware	Camera and computer required

TABLE VIII

required						
Limitations	Needs understanding	shared	Calibration use	and glove	Lighting, occlusion, dataset limits	

VI. LIMITATIONS AND FUTURE SCOPE

A. Limitations • Recognition may be unreliable under poor lighting, occlusion, cluttered backgrounds, or unusual camera angles. • A model trained on a limited dataset may not generalize to new signers or regional variations. • A static-pose classifier cannot fully recognize signs that depend on motion, facial expression, or body posture. • The system recognizes only the labels represented in its training data and should not be described as a complete sign-language translator. • Performance depends on the dataset, camera, hardware, preprocessing, model, and evaluation protocol. B. Future Scope • Expand the dataset with diverse signers and balanced examples for each class. • Add temporal models for motion-based signs and sequence recognition. • Explore speech output, sentence assembly, and multilingual interface options. • Improve robustness to backgrounds, lighting, and partial hand occlusion. • Evaluate on independent users and report per-class metrics and latency. • Extend recognition to two-hand gestures, facial cues, and broader sign vocabularies where appropriate.

VII. CONCLUSION

The proposed Real-Time Sign Language Recognition System combines camera input, hand-landmark detection, feature extraction, classification, and text output in a modular pipeline. It offers a practical foundation for recognizing a defined set of signs with accessible, low-cost hardware. Its usefulness depends on careful dataset design, real-time implementation, and honest evaluation. Before submission, the system should be implemented and tested, and measured results should replace placeholders. The paper should clearly describe the supported sign vocabulary, dataset, training procedure, hardware, limitations, and evaluation protocol. Future work may extend the prototype to dynamic signs, broader vocabularies, and speech output.

REFERENCES

[1] O. M. S. Al-Qatab and A. M. H. Al-Hamadi, "A Survey on Sign Language Recognition Systems," (verify full bibliographic details and use an authoritative source before submission).

[2] Google, "MediaPipe Hands: On-device Real-time Hand Tracking," MediaPipe documentation and associated research publications. Verify the exact citation details for the version used.

[3] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.

[4] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, pp. 273–297, 1995.

[5] M. A. Nielsen, Neural Networks and Deep Learning, online book, 2015. Cite only if used in the methodology.

[6] Add the exact dataset paper, dataset repository, and model/library documentation actually used in the project; verify all bibliographic details before submission. STUDENT CHECKLIST BEFORE SUBMISSION • Replace the sample author and guide details if spelling or initials need correction. • Replace the evaluation plan with actual results from the implemented system. • Include the student's own system screenshots, diagrams, dataset details, and methodology. • Verify references and cite the dataset and software libraries actually used. • Check similarity, permissions, dataset licensing, and privacy/consent requirements.