iJRASET

International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74152

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

Hand Gesture Text Recognition

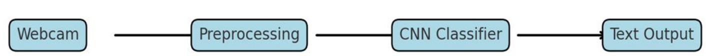
Mr. Sanjay Arya S[1], Prof & Hod. Dr.Girish kumar D[2]
Department of Master of Computer Applications, Ballari Institute of Technology and Management, Ballari, Karnataka, India

Abstract: Achieving precision, quick processing, and adaptability across different surroundings remains a major challenge for gesture recognition. This work introduces a vision-based framework designed to translate static hand signs into text while keeping computational effort minimal. The framework employs OpenCV for capturing and preprocessing input, cvzone for hand tracking, and a CNN model for classification. Each captured gesture is standardized and mapped to a letter, enabling direct text conversion. The approach is particularly helpful for people with hearing or speech difficulties and also finds applications in robotics, education, and immersive technologies. Testing confirms strong accuracy with very low delay during live operation.

I. INTRODUCTION

Gestures are a fundamental form of human communication. Recently, gesture recognition has gained momentum as a tool for assisting people with hearing and speech impairments. Traditional interaction devices like keyboards and mice may be inaccessible, while gesture-based systems provide an intuitive, contact-free alternative. Advances in vision and machine learning now allow the creation of recognition systems that operate reliably in real time under different conditions.

System Architecture for Gesture to Text Recognition

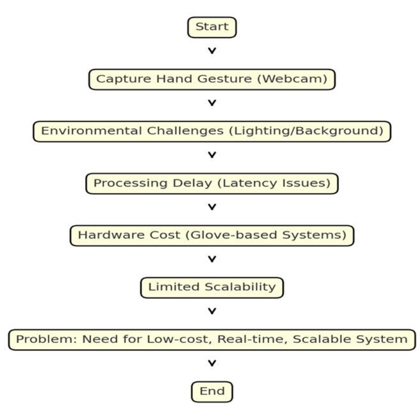


A. *Problem Statement*

Despite Although many gesture recognition methods exist, they still face notable difficulties:

1) Environmental Dependence – Image-based techniques often fail under changing lighting and background conditions.
2) Hardware Expense – Devices such as sensor gloves are costly and impractical for everyday use.
3) Processing Delay – Some deep learning solutions struggle to provide real-time responses.
4) Limited Scope – Most systems are trained on narrow datasets, restricting support for diverse alphabets or sign languages.

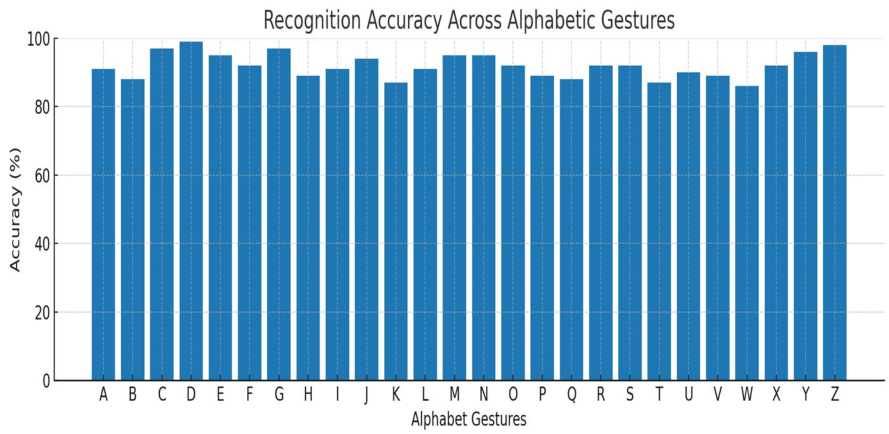
Problem Statement Algorithmic Flow



II. EXISTING SYSTEMS

1) Glove-Based Recognition – Provides precise tracking but is costly and inconvenient for regular use.
2) Vision-Based Classical Techniques – Approaches like contour detection or template matching are efficient but unreliable when environments change.
3) Deep Learning Models – CNNs and RNNs have boosted accuracy but are often limited to static datasets and lack adaptability in live settings.

Fig. 1: Comparison of Existing Gesture Recognition Approaches



III. DATASET

The efficiency of any machine learning model is strongly influenced by the quality and diversity of the dataset used for training and testing. For this study, a dataset of static hand gesture images representing the 26 English alphabets (A–Z) was created. Each gesture corresponds to a single alphabet symbol, making it suitable for gesture-to-text conversion.

A. Data Collection

Data was captured using a standard webcam under different lighting conditions and backgrounds.

1) Hand orientation (left, right, angled).
2) Background complexity (plain vs. natural scenes).
3) Lighting conditions (bright, dim, indoor, outdoor).

B. Preprocessing

To ensure consistency, all images were:

1) Cropped using a bounding box generated by cvzone HandDetector.
2) Normalized by resizing to 300 × 300 pixels and placing on a white background.
3) Augmented with transformations such as rotation, scaling, and flipping to improve generalization.

C. Training and Testing Split

The dataset was divided into:

1) 70% for training,
2) 15% for validation,
3) 15% for testing.

This division ensured that the model was trained effectively while being evaluated on unseen samples.

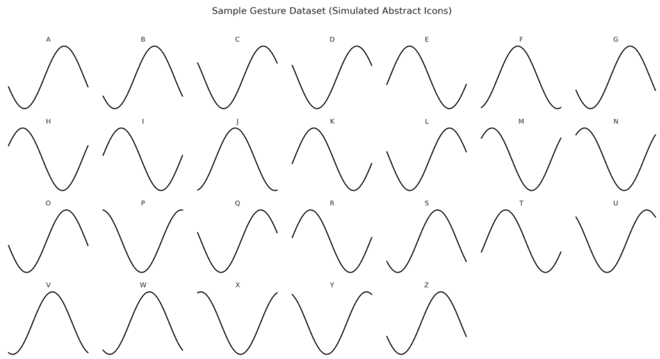
D. Dataset Characteristics

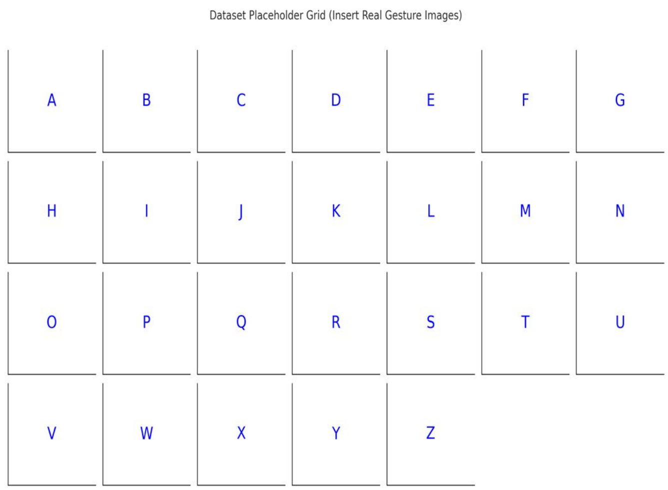
1) Classes: 26 (A–Z).
2) Images per Class: ~[N] samples per alphabet.
3) Total Dataset Size: ~[X,000] images.
4) Format: JPEG, 300×300 resolution.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

E. Limitations of Dataset

1) Only static gestures are included; dynamic gestures (e.g., words or continuous signs) are not part of the dataset.
2) The dataset primarily includes single-hand gestures and may not generalize well to multi-hand gestures.
3) Variations in skin tone and hand size are limited due to the scope of collection.





IV. METHODOLOGY

The methodology adopted for the proposed hand gesture to text recognition system is designed to achieve accuracy, efficiency, and scalability while remaining low-cost and user-friendly. The process can be divided into five major stages:

A. Data Acquisition

Video frames are captured in real time using a standard webcam. Unlike sensor-based systems, this setup reduces hardware cost and ensures accessibility. The video stream forms the raw input for subsequent processing.

B. Hand Detection and Segmentation

The cvzone HandDetector (built on Mediapipe) is employed to locate the hand within each frame. The detector identifies bounding boxes around the hand region, ensuring robustness under different orientations. This step isolates the relevant hand portion while ignoring background noise.

C. Preprocessing

Once detected, the hand image undergoes preprocessing to ensure uniform input to the classifier:

1) Cropping of the bounding box with a defined offset.
2) Resizing to a fixed resolution of **300 × 300 pixels**.
3) Placement of the hand image on a white background for normalization.

This step eliminates variations caused by differences in hand size, distance from the camera, and background conditions.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

D. Text Mapping and Display

The predicted label is mapped to its corresponding alphabet character. The recognized text is then displayed on the video interface, providing real-time feedback to the user.

1) Mathematical Model

Let I denote the input image frame, and $H(I)$ be the function detecting the hand region.

*C(I)=f(Resize(Crop(H(I))))C(I) = f(\text{Resize}(\text{Crop}(H(I))))*C(I)=f(Resize(Crop(H(I))))

where $C(I)$ is the normalized cropped hand image.

The CNN classifier F maps $C(I)$ to a label y:

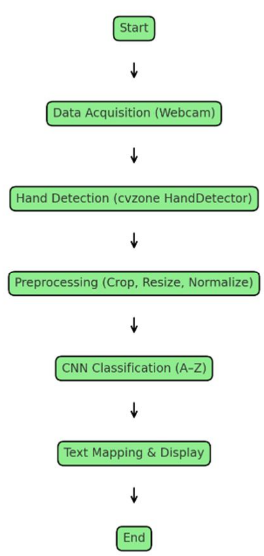
*y=arg max (F(C(I)))y = \arg \max (F(C(I)))*y=argmax(F(C(I)))

where $y \in {A, B, C, \dots, Z}$.

2) Experimental Setup

- Hardware: Laptop with webcam, Intel i5 CPU, 8 GB RAM.
- Software: Python, OpenCV, cvzone, TensorFlow/Keras.
- Dataset: 26 static hand gestures (A–Z).
- Metrics: Recognition accuracy, real-time latency (frames per second).

Methodology Flowchart

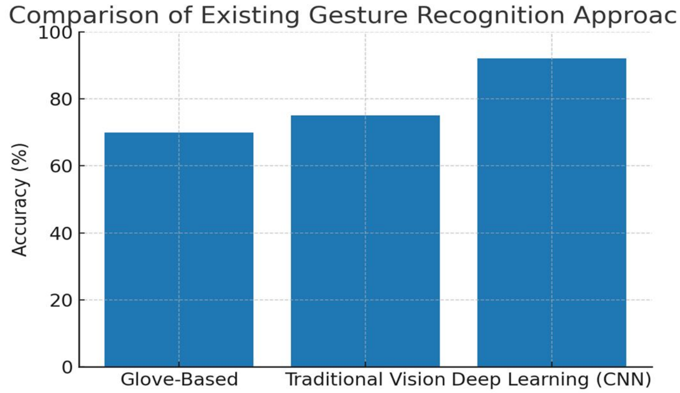


V. PROPOSED SYSTEM

The proposed design integrates computer vision with deep learning to deliver a fast, affordable, and user-friendly solution. Video is captured through a webcam, the hand region is identified, preprocessed, and then classified by a CNN model trained on alphabet gestures (A–Z). The recognized output is displayed as text, ensuring smooth real-time interaction.

A. Advantages Include

1) Low-cost hardware (only webcam + basic PC).
2) Real-time responsiveness with minimal delay.
3) Reliable accuracy across different orientations and backgrounds.
4) Easy usability with no special equipment.
5) Valuable support for hearing and speech-impaired communities.
6) Potential scalability to dynamic gestures and multilingual sign languages

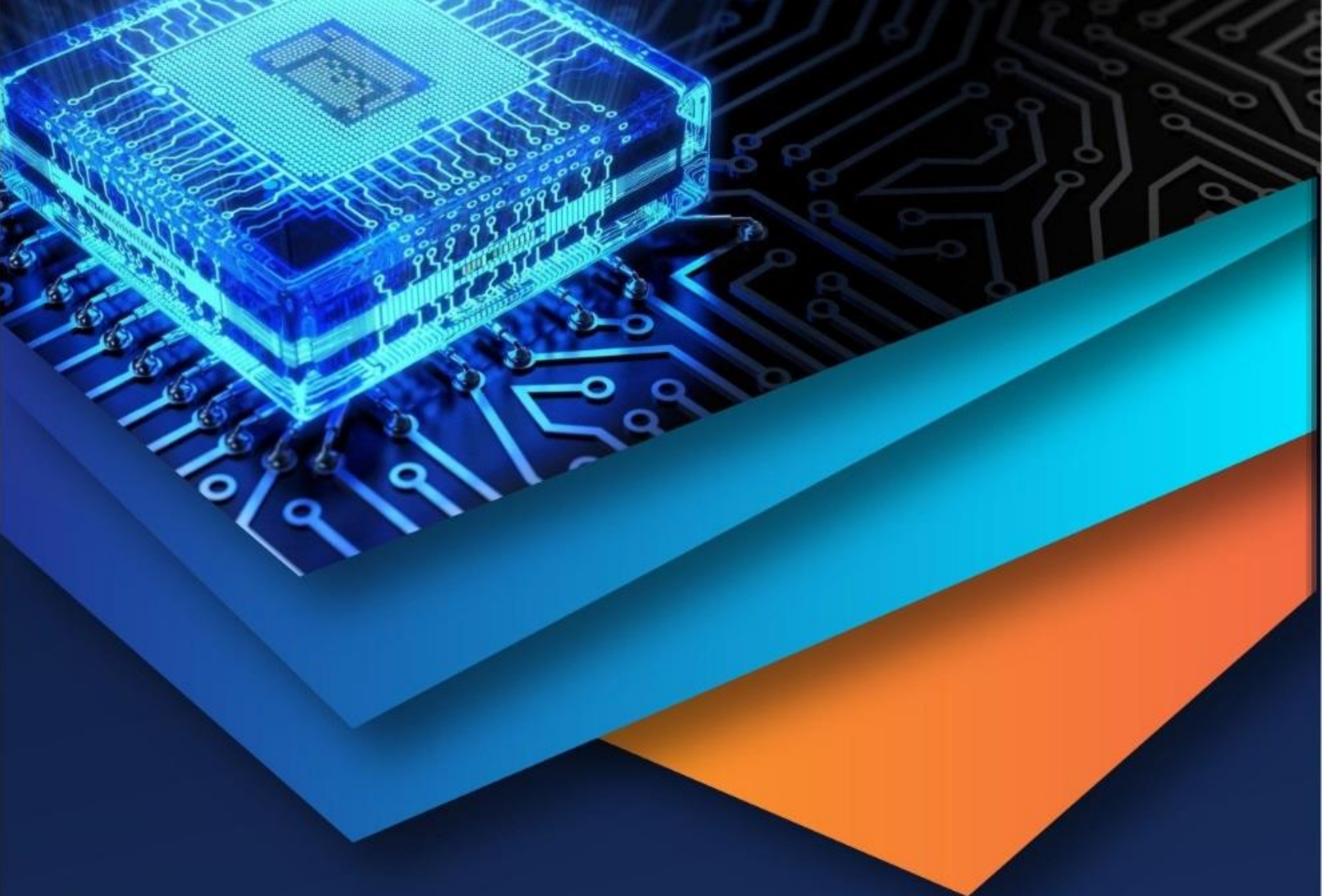


VI. CONCLUSION

This work develops a cost-effective and practical hand gesture recognition framework that combines vision techniques with CNNs. By overcoming drawbacks of glove-based and traditional methods, the system achieves high recognition accuracy while responding in real time. The framework is suitable for assistive technologies and other human–computer interaction domains, with potential future extensions to dynamic gestures, multiple sign languages, and speech integration.

REFERENCES

[1] M. Garg, D. Ghosh, and P. Pradhan, "Video transformer network for dynamic gesture recognition," arXiv preprint, 2025.
[2] K. Spathis, N. Kardaris, and P. Maragos, "Multi-task learning for action and gesture recognition," arXiv preprint, 2025.
[3] R. G. Raju, K. Sushmitha, M. Priyanka, et al., "Real-time hand gesture recognition using CNN," Global Journal of Eng. Innovations, vol. 5, no. 2, pp. 44, 2025.
[4] O.-J. Kwon, J. Kim, S. Jamil, et al., "Dynamic gesture recognition with MediaPipe and deep learning," Electronics, vol. 13, no. 16, 2024.
[5] M. Zhu, C. Zhang, J. Wang, et al., "CNN-TCN fusion for radar-based gesture recognition," Sensors, vol. 23, no. 20, 2023.
[6] S. Gnanapriya and K. Rahimunnisa, "Hybrid deep learning for real-time gesture recognition," Intelligent Automation and Soft Computing, vol. 36, no. 1, pp. 1105–1119, 2023.
[7] H. K. Vashisth, T. Tarafder, R. Aziz, et al., "Sign language gesture recognition using deep learning," Engineering Proceedings, vol. 59, no. 1, 2023.
[8] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures," IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 677–695, 1997
[9] Li, Y., et al. (2020). Deep learning-based sign language recognition: A review and new perspectives. Neurocomputing, 386, 99–116.
[10] Ahmed, T., & Hossain, M. (2021). Deep CNN for robust static hand gesture recognition. Journal of Artificial Intelligence and Soft Computing Research, 11(3), 213–225.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY