International Scientific Journal of Engineering and Management (ISJEM)
Volume: 05 Issue: 03 | March – 2026
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata
ISSN: 2583-6129
DOI:10.55041/ISJEM05608

Real-Time Sign Language Recognition System using Computer Vision and Deep Learning

Akshay B. Nagde
,akshaynagade97@gmail.com

Dr. M.V. Lande
milind.jdiet@gmail.com

Pratham C Durge
durgecp@gmail.com

Om D. Chavan
omc75243@gmail.com

Pranav S. Shelke
pranavshelke8975@gmail.com

Sanskruti C. Shrirang
sanskrutishrirang@gmail.com

Department Of Computer Engineering Government Polytechnic Gadchiroli, Maharashtra, India

---***---

Abstract:

Sign language serves as a communication method for individuals who are deaf or mute. Despite this, there are still communication challenges between those with hearing impairments and those without, primarily due to the absence of automated interpretation systems. This project introduces a real-time sign language recognition system that leverages computer vision and machine learning. The system utilizes a live camera feed to capture hand gestures, identifies hand landmarks with Media Pipe, and employs a trained Convolutional Neural Network (CNN) model to classify these gestures. The identified gestures are then translated into text, with suggestions for commonly used words provided to enhance communication efficiency. The system is designed to be user-friendly, affordable, and capable of operating in real time without the need for specialized hardware, thus aiding in closing the communication gap and fostering inclusivity.

Key Words:

Sign Language Recognition, Hand Gesture Recognition, Machine Learning, Deep Learning, CNN, Computer Vision, Media Pipe

1.INTRODUCTION

Communication is crucial in human interactions, allowing people to exchange ideas, feelings, and information. Deaf and mute individuals rely on sign language to communicate. However, a significant issue arises because most people do not understand sign language, leading to a considerable communication gap between those who can hear and the general population. This gap often restricts social interactions, educational opportunities, and access to vital services for deaf and mute individuals.Traditionally, interacting with individuals who are deaf and mute necessitates either a human interpreter or familiarity with sign language. These options are not always feasible, cost-effective, or accessible in real-time scenarios. Consequently, there is a significant demand for an automated system capable of translating sign language gestures into a format that is easily comprehensible to everyone.

Recent progress in computer vision and machine learning has enabled the creation of intelligent systems that can interpret visual patterns. Specifically, deep learning methods like Convolutional Neural Networks (CNNs) have demonstrated outstanding results in tasks involving image and gesture recognition. These technologies allow machines to learn intricate patterns from data and make precise predictions without the need for manual rule-based programming.

This project presents a real-time sign language recognition system developed using computer vision and machine learning methods. The system captures hand gestures through a camera and utilizes MediaPipe technology to accurately detect hand landmarks. Based on these detected gestures, the system identifies the corresponding signs and displays the output as text instantly, improving communication accessibility.

The system improves user convenience by providing suggestions of frequently used words according to the recognized gestures. This capability allows users to construct meaningful phrases and sentences in a quicker and more efficient manner. The proposed approach is affordable, simple to use, and operates without the need for specialized hardware, making it a practical solution for real- world assistive applications and promoting accessible communication.

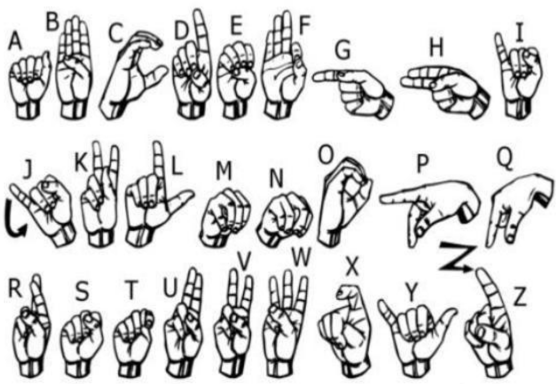
International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05608
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

2.LITERATURE REVIEW

Sr. No.	Authors	Purpose	Methodology	Improvement	Limitations
1	Sharma & Chaudhary (2021)	Sign language translation	Image processing with Machine Learning	Improved accessibility for deaf users	Lighting and background variations
2	Kumar et al. (2023)	Vision-based gesture recognition	Deep learning models	High recognition precision	High computational cost
3	Joshi & Murali (2025)	Gesture recognition	CNN-based classification	Improved real-time accuracy	Requires large and diverse dataset
4	Patel et al. (2022)	Real-time hand gesture recognition	CNN with image preprocessing	Faster response time	Sensitive to lighting conditions
5	Singh & Verma (2020)	Static sign language recognition	Image processing techniques	Simple implementation	Lower accuracy compared to deep learning
6	Rao et al. (2021)	Hand gesture classification	Machine learning classifiers	Reduced training complexity	Limited scalability

3.SYSTEM OVERVIEW

The given image represents the alphabet-level hand gestures of sign language, where each hand pose corresponds to a specific English alphabet letter from A to Z. These gestures are commonly associated with American Sign Language (ASL) and are widely used as a foundational learning set in sign



language recognition systems. Each gesture is defined by a unique combination of finger positions, hand orientation, and shape, making it suitable for visual- based recognition.

This image serves as an essential visual reference in sign language recognition systems because it clearly represents static hand gestures. Static gestures refer to hand positions that remain unchanged for a brief period, unlike dynamic gestures that involve continuous movement. Most sign language alphabet gestures are static, which makes them easier to capture, annotate, and classify using computer vision and machine learning approaches.

From a machine learning viewpoint, hand gesture images function as training data or reference samples during model development. While preparing the dataset, several images of each alphabet gesture are collected under varying lighting conditions, backgrounds, and viewing angles. These images are then utilized to train deep learning models, such as Convolutional Neural Networks (CNNs), enabling them to learn unique features including finger positions, palm orientation, finger curvature, and spatial relationships between fingers.

In real-time sign language recognition systems, like the one proposed in this project, hand gestures are captured using a camera and processed continuously frame by frame. With the help of tools such as MediaPipe, important hand landmarks are detected and extracted. These landmarks are then compared with the patterns learned from the alphabet gesture dataset. Based on the detected hand pose, the trained model predicts the most likely alphabet, which is subsequently converted into readable text for the user.

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05608
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

Overall, this image serves as a conceptual and visual foundation for sign language recognition systems. It helps in understanding the mapping between hand gestures and textual representation, which is essential for building assistive technologies aimed at improving communication for deaf and mute individuals.

4.HARDWARE REQUIREMENTS

Webcam or Laptop Camera

A webcam integrated into a laptop functions as an imaging device that records live Image or video data. In the proposed system for sign language recognition, the camera captures the user's hand movements continuously. This real-time visual input acts as the initial source of data for detecting and recognizing gestures.

The use of a laptop camera eliminates the need for external hardware such as sensors or wearable devices. This makes the system more accessible, economical, and easy to deploy on commonly available devices. The camera's ability to provide consistent input ensures reliable interaction between the user and the system.

Uses of Laptop Camera

- Capturing real-time hand gestures
- Providing visual input for computer vision systems
- Supporting gesture-based recognition applications
- Enabling assistive communication technologies Laptop

A laptop is a portable computing device that serves as the central processing unit of the proposed system. It performs all major operations including video processing, feature extraction, machine learning computation, and result visualization. In this project, the laptop hosts the software environment required for sign language recognition.

The system utilizes a laptop to execute programming libraries such as Python, OpenCV, MediaPipe, and TensorFlow for processing the camera input and operating the trained convolutional neural network model. The laptop carries out tasks including hand landmark detection, gesture recognition, and real-time text generation. Furthermore, it offers a graphical user interface that allows users to view the recognized results directly on the screen.

Because of its portability and strong processing power, the laptop enables the system to operate in different environments such as classrooms, homes, and public areas. By combining both hardware and software resources, the laptop plays a vital role in implementing a real-time and user-friendly sign language recognition system.

Uses of Laptop

- Processing real-time video data
- Running machine learning and deep learning models
- Executing computer vision algorithms
- Displaying recognized text and system output

5.SOFTWARE REQUIREMENTS

- Python
- OpenCV
- MediaPipe
- TensorFlow / Keras
- Tkinter (for GUI)

6.WORKING OF THE SYSTEM

6.1 Sign Language recognition Architecture

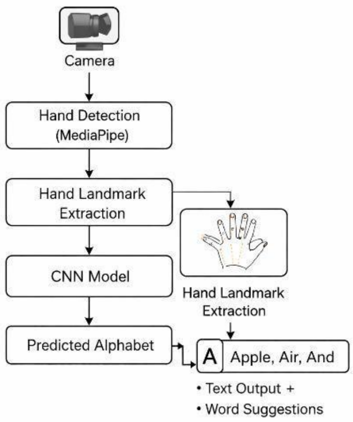


Fig. 3.1 System Architecture of the Proposed Sign Language Recognition on

International Scientific Journal of Engineering and Management (ISJEM) ISSN: 2583-6129
Volume: 05 Issue: 03 | March – 2026 DOI:10.55041/ISJEM05608
An International Scholarly || Multidisciplinary || Open Access || Indexing in all major Database & Metadata

The developed sign language recognition system is designed to interpret hand gestures and transform them into textual output using computer vision and machine learning approaches. According to the system architecture, the first step involves obtaining real-time video input through a regular camera.

After capturing the video frames, the system performs hand detection using the MediaPipe framework. MediaPipe identifies the hand in each frame and extracts important landmarks such as finger joints and palm positions. These landmarks create a structured representation of the hand gesture, which helps minimize the impact of background complexity and lighting variations compared to conventional image- based methods.

After hand detection, the extracted landmark features are sent to a Convolutional Neural Network (CNN) model for classification. The CNN is trained on a labeled dataset of sign language gestures and learns spatial features such as hand shape, finger orientation, and the relative positions of landmarks. Using these learned patterns, the model predicts the corresponding sign language alphabet.

Once the gesture is classified, the predicted alphabet is transmitted to the system's output module and displayed as text on the graphical user interface in real time. The system also provides suggestions of frequently used words related to the predicted alphabet, helping users create meaningful words and sentences efficiently.

The system architecture ensures efficient data processing from gesture capture to text output by integrating MediaPipe-based hand landmark extraction with a CNN-based model for real-time gesture classification and text generation.

6.2 CNN Architecture

The figure shows the Convolutional Neural Network (CNN) architecture used to classify sign language gestures. The process begins with a hand gesture image captured by a camera, which is resized and formatted before being provided to the CNN model. This image represents a static gesture, such as a sign language alphabet, and serves as the input for feature extraction.

The image is then processed through convolutional layers with small filters (e.g., 3×3) and ReLU activation functions to extract important features like edges,contours.

As the data moves deeper into the network, the extracted features become more meaningful and help distinguish between similar hand gestures.

6.2 CNN Architecture

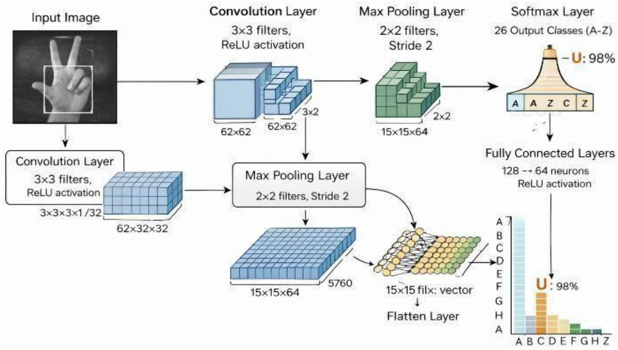


Fig. 8.1 CNN Architecture Used for Sign Language Gesture Classification

Following the convolution layers, max pooling layers are applied to reduce the spatial dimensions of the feature maps. Pooling helps in retaining the most important features while reducing computational complexity and sensitivity to small variations in hand position. The pooled feature maps are then flattened into a one- dimensional feature vector, which serves as input to the fully connected layers.

Finally, the fully connected layers perform classification based on the learned features. The soft max output layer computes probabilities for each sign language class (A– Z), and the class with the highest probability is selected as the predicted output. As shown in the figure, the network successfully predicts the gesture with a high confidence score. This CNN-based approach enables accurate and efficient real-time sign language recognition.

7.MACHINE LEARNING METHODOLOGY

The proposed system applies a Convolutional Neural Network (CNN) to recognize sign language gestures accurately. As illustrated in the CNN workflow diagram, the process begins after hand landmark extraction, where structured feature data is provided to the CNN model to learn spatial patterns automatically.

The input data passes through convolutional layers where multiple filters extract key features such as edges, finger orientation, and the relative positions of hand landmarks, allowing the system to distinguish between similar sign language alphabets.

The extracted high-level features are then forwarded to fully connected layers, where these features are combined to perform gesture classification. The final output layer applies a softmax activation function to determine the probability of each sign language class. The class with the highest probability is chosen as the predicted alphabet or gesture.

During the training phase, the CNN model is trained on a labeled dataset of hand gesture images using TensorFlow and Keras frameworks. The trained model is saved in .h5 format and later utilized for real-time prediction. This CNN-based approach provides accurate, fast, and reliable recognition of sign language gestures, making the system suitable for real-time assistive communication applications.

8.FEATURES OF THE SYSTEM

The proposed sign language recognition system is developed to deliver accurate and real-time gesture recognition using machine learning techniques. Its real- time processing capability allows gestures to be detected instantly, providing smooth and natural interaction for users.

A notable feature of the system is the integration of daily-used word suggestions. After identifying individual alphabets, the system recommends frequently used words, helping users form meaningful words and sentences more efficiently.

Furthermore, the system is designed to be simple and accessible. It operates using only a standard laptop camera and does not require specialized hardware or wearable devices. Due to its portability and low cost, the system is well suited for real-world assistive communication applications.

9.ADVANTAGES

- Enables effective communication between deaf and mute individuals and the general population.
- Provides real-time sign language recognition with fast response time.
- Reduces dependency on human interpreters
- Uses a cost-effective setup with only a laptop and inbuilt camera.
- Does not require any special sensors, gloves, or wearable devices.
- Machine learning (CNN)-based model provides accurate and reliable gesture recognition.
- User-friendly graphical interface makes the system easy to use.

10. LIMITATIONS

- System performance may vary depending on lighting conditions.
- The current system supports a limited set of gestures.
- Accuracy may decrease when multiple hands appear in the frame.
- Complex or cluttered backgrounds can affect detection accuracy.

11. APPLICATIONS

The sign language recognition system can be utilized in several real-world applications. It can act as a communication support tool for individuals who are deaf or mute during everyday interactions. In educational settings, the system may be integrated into smart classrooms to encourage inclusive learning environments. It also contributes to human–computer interaction by allowing gesture-based control. Additionally, the system can be implemented in public service centers, hospitals, and customer support platforms to improve accessibility and inclusiveness

12. CONCLUSION

The real-time sign language recognition system demonstrates the effective application of machine learning and computer vision to solve real-world communication challenges. By combining MediaPipe-based hand landmark detection with a CNN-based deep learning model, the system provides accurate and reliable real-time gesture recognition.

The inclusion of daily-used word suggestions enhances usability and reduces communication effort. Although environmental conditions and dataset size can affect performance, the system establishes a strong foundation for future improvements. Overall, this project proves that machine learning-driven assistive technologies can significantly improve accessibility and contribute to building an inclusive and technology-enabled society.

for communication.

13. RESULTS AND PERFORMANCE ANALYSIS

The performance of the proposed system was assessed through real-time testing with various hand gestures under different lighting and background conditions. The trained CNN model achieved satisfactory accuracy in identifying commonly used sign language alphabets. During live operation, the system exhibited low latency, allowing smooth interaction without noticeable delays. These findings suggest that the system is reliable and suitable for real-world applications.

ISSN: 2583-6129
DOI:10.55041/ISJEM05608

REFERENCES

- Joshi, R., & Murali, S., "Real-Time Hand Gesture Recognition Using Convolutional Neural Networks," *International Journal of Computer Vision and Machine Learning*, vol. 12, no. 3, pp. 45–52, 2025.
- Sharma, P., & Chaudhary, A., "Sign Language Translation System Using Image Processing and Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 10, no. 6, pp. 210–215, 2021.
- Kumar, A., Singh, R., & Patel, M., "Vision-Based Hand Gesture Recognition Using Deep Learning," *International Journal of Advanced Research in Computer Science*, vol. 14, no. 2, pp. 98–104, 2023.
- Patel, N., Shah, K., & Mehta, D., "Real-Time Hand Gesture Recognition System Using CNN," *International Conference on Artificial Intelligence and Smart Systems*, pp. 112–117, 2022.
- Singh, V., & Verma, P., "Static Sign Language Recognition Using Image Processing Techniques," *International Journal of Signal Processing Systems*, vol. 8, no. 1, pp. 25–30, 2020.
- Rao, S., Kulkarni, A., & Deshmukh, P., "Machine Learning-Based Hand Gesture Classification," *International Journal of Emerging Technologies in Engineering Research*, vol. 9, no. 4, pp. 56–61, 2021.
- Zhang, X., Yang, L., & Chen, Y., "Deep Learning Approaches for Hand Gesture Recognition," *IEEE Access*, vol. 9, pp. 123456–123465, 2021.
- Google, "MediaPipe Hands: Real-Time Hand Tracking," Google Research Documentation, 2022.
- Goodfellow, I., Bengio, Y., & Courville, A., *Deep Learning*, MIT Press, 2016.
- Chollet, F., *Deep Learning with Python*, Manning Publications, 2018.