

Sign Language Recognition using MediaPipe and Machine Learning

¹Prof. Dr. B. Geetha Vani, ²K. Chandrakanth, ³S. Anusha, ⁴B. Harish

¹Professor in Dept. of CSE, G. Pulla Reddy Engineering College, Kurnool, Andhra Pradesh, India,

^{2,3,4}Undergraduate student in Dept. of CSE, G. Pulla Reddy Engineering College, Kurnool, Andhra Pradesh, India

Abstract - Sign language recognition systems play a vital role in improving communication for hearing and speech-impaired individuals; however, real-time gesture recognition is affected by variations in hand positioning, background noise, and lighting conditions. In this work, MediaPipe is used for hand landmark detection along with a Random Forest classifier for real-time gesture recognition. A workflow is designed to collect data, extract normalized landmarks, and train the model. The system is deployed through a web-based interface that converts hand signs representing alphabets (A-Z) and digits (0-9) into text with word and sentence formation. Experimental results show reliable performance and improved accuracy, demonstrating the suitability of the proposed system for assistive communication applications.

Key Words: Sign Language recognition, hand gesture detection, MediaPipe, Random Forest, real-time system, assistive communication

1. Introduction

Sign language recognition systems are widely used to assist communication for hearing and speech-impaired individuals in various real-world environments. However, real-time hand gesture recognition systems often suffer from variations in hand positioning, background noise, and poor lighting conditions. These limitations arise due to camera quality, environmental factors, and diverse hand orientations. As a result, the accuracy of gesture interpretation is frequently insufficient for reliable communication, highlighting the need for effective gesture recognition techniques.

Several approaches have been explored to improve the performance of gesture recognition systems, ranging from traditional image processing techniques to machine learning and deep learning-based methods. In recent years, hand tracking frameworks combined with machine learning algorithms have shown significant improvements over conventional approaches. In particular, landmark-based detection methods have gained attention due to their ability to capture precise hand features. Prior studies have demonstrated the effectiveness of such approaches for recognizing hand gestures, while also highlighting challenges related to dataset variability, lighting conditions, and computational efficiency.

Recent advancements in hand tracking technologies have led to the development of efficient frameworks capable of detecting detailed hand landmarks in real time. Among these, MediaPipe has gained attention for its ability to accurately detect and track hand movements under varying conditions. Unlike traditional methods that rely heavily on raw images, MediaPipe extracts structured landmark features, making the system more robust and efficient. Motivated by these capabilities, this work develops a real-time sign language recognition system using MediaPipe for feature extraction and a Random Forest classifier for gesture classification. The system is deployed through a web-based interface, focusing on real-time performance and practical usability rather than complex model architectures.

2. Related Work

Sign language recognition and hand gesture detection have been extensively studied to improve communication systems for hearing and speech-impaired individuals. Early approaches relied on traditional image processing techniques such as skin color detection, contour extraction, and thresholding; however, these methods often failed to accurately capture complex hand gestures under varying lighting and background conditions. With the advancement of machine learning, landmark-based and feature extraction techniques have demonstrated significant improvements in gesture recognition tasks by learning patterns from structured hand representations.

More recently, deep learning-based methods such as Convolutional Neural Networks (CNNs) have been introduced for hand gesture recognition due to their ability to automatically learn spatial features from images. CNN-based models have shown promising results in recognizing complex gestures and improving classification accuracy. Several studies have explored the use of deep learning for sign language recognition, highlighting their effectiveness in handling large-scale datasets while also noting challenges related to high computational cost, large training data requirements, and real-time implementation constraints.

In the context of real-time gesture recognition, existing works have primarily focused on training complex deep learning models or using hybrid approaches combining image processing and neural networks.

In contrast, this work emphasizes a lightweight and practical approach by utilizing MediaPipe for efficient hand landmark detection and a Random Forest classifier for gesture classification.

While existing studies have demonstrated the effectiveness of deep learning models for gesture recognition, most approaches focus on improving accuracy through complex architectures or large datasets. These methods often require significant computational resources and are not always suitable for real-time deployment. In contrast, the proposed work adopts an application-oriented approach by directly using normalized hand landmarks with a machine learning classifier, enabling efficient real-time recognition without heavy computational requirements.

3. Methodology

The overall workflow begins with the acquisition of hand gesture images captured using a webcam for different classes representing alphabets (A-Z) and digits (0-9). These images represent real-time conditions, including variations in hand positioning, background noise, and illumination differences. Since the data is collected from real-world environments, it reflects practical challenges encountered in gesture recognition systems.

After acquisition, each image undergoes processing using MediaPipe Hands before being passed to the classification model. The processing stage includes detection of hand landmarks and extraction of normalized coordinate features by subtracting minimum x and y values. This normalization ensures consistency across different hand positions and scales, improving the robustness of the system under varying conditions.

The feature representation is constructed using the extracted hand landmarks, where each gesture is represented as a set of normalized (x, y) coordinates. These features are used to train a Random Forest classifier, which learns patterns corresponding to different gestures. The trained model is capable of performing real-time classification and predicting the corresponding alphabet or digit based on the detected hand gesture.

To evaluate enhancement performance, the dataset is divided into training and testing sets using a standard split. The model performance is assessed using metrics such as accuracy, precision, recall, and F1-score. Additionally, the system is deployed through a web-based interface that enables real-time gesture recognition.

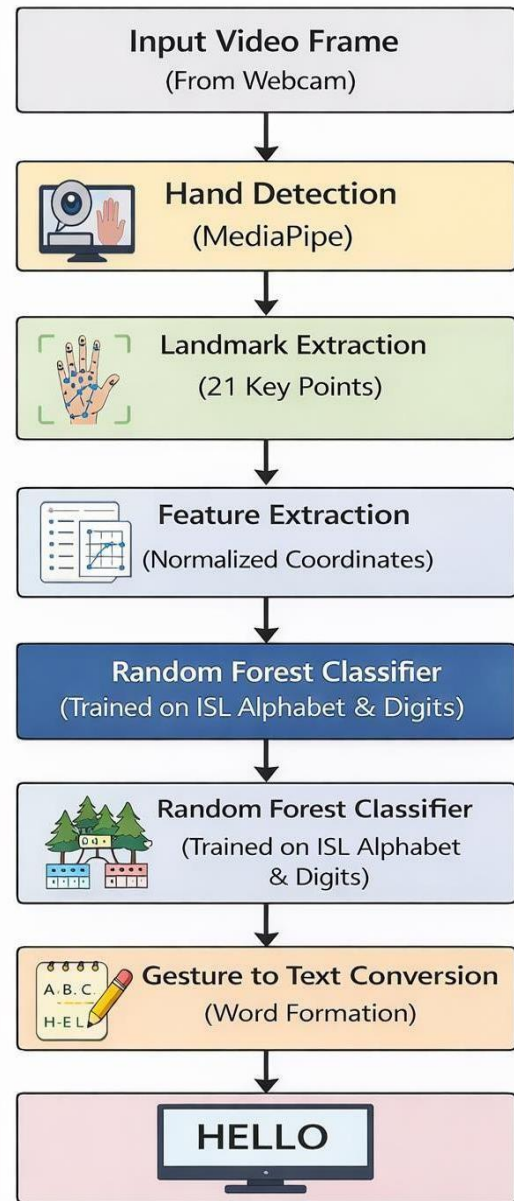


Figure - 1: Pipeline of Gesture Recognition

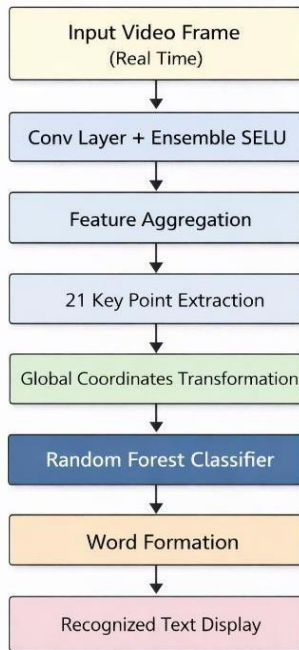


Figure - 2 : Architecture of Gesture Recognition

4. Experimental Setup

The experimental evaluation of the proposed sign language recognition system was conducted using a collection of hand gesture images captured through a webcam. The dataset was designed to reflect real-time conditions rather than controlled laboratory environments. A set of images was collected for each class representing alphabets (A-Z) and digits (0-9). The data collection process involved capturing multiple samples per class under different orientations and lighting conditions to ensure variability and robustness.

The dataset includes a wide range of variations such as different hand positions, background conditions, and illumination levels, which are commonly encountered in real-world scenarios. It also incorporates variations in gesture orientation and scale, which can affect recognition accuracy.

The model architecture utilizes a landmark-based feature extraction framework using MediaPipe Hands for real-time gesture detection. In this approach, the hand tracking module focuses on identifying 21 key landmark points representing finger joints and palm structure. These landmarks capture meaningful spatial relationships of the hand. The normalization process plays a crucial role in reducing variations caused by hand position and scale differences, enabling consistent feature representation across different inputs.

To further utilizes a landmark-based feature extraction framework using MediaPipe Hands for real-time gesture detection. In this approach, the hand tracking module focuses on identifying 21 key landmark points representing finger joints and palm structure. These landmarks capture meaningful spatial relationships of the hand. The normalization process plays a crucial role in reducing variations caused by hand position and scale differences, enabling consistent feature representation across different inputs.

Each gesture sample in the dataset is processed individually through the detection and classification pipeline. The system records several evaluation parameters including accuracy, precision, recall, and F1-score to measure the effectiveness of the recognition process. In addition to quantitative evaluation, real-time testing is performed using the web-based interface, where users interact with the system by performing gestures. This evaluation allows assessment of the model's performance under practical conditions and demonstrates its effectiveness compared to traditional gesture recognition approaches.

5. Evaluation Metrics

Due to the absence of standardized benchmark datasets for real-time hand gesture recognition under varying conditions, evaluation accuracy is computed using classification-based performance metrics. The effectiveness of the proposed system is assessed by comparing predicted gesture labels with actual class labels obtained from the dataset.

Accuracy is calculated as the ratio of correctly predicted gestures to the total number of samples. It provides an overall measure of the model's performance in correctly classifying hand gestures. However, accuracy alone may not fully represent performance when class distributions vary, making additional metrics necessary for comprehensive evaluation.

Similarly, precision and recall are computed to measure the model's ability to correctly identify gesture classes. Precision evaluates how many predicted gestures are actually correct, while recall measures how effectively the model identifies all relevant gesture instances. These metrics provide insight into classification reliability under different conditions.

In addition to accuracy, precision, and recall, the F1-score is calculated as the harmonic mean of precision and recall. This metric provides a balanced evaluation of the model's performance, especially when dealing with variations in gesture recognition. These evaluation metrics collectively enable effective assessment of the proposed system under real-time conditions.

6. Result and Analysis

The experimental results demonstrate noticeable improvements in real-time hand gesture recognition accuracy and prediction consistency. The proposed system effectively identifies hand gestures representing alphabets and digits under varying conditions. Compared to traditional image-based approaches, the landmark-based method provides stable predictions and reduces errors caused by background noise and lighting variations.

Hand regions and finger positions are accurately detected using MediaPipe, enabling precise classification of gestures. In several cases, gestures that are visually similar are correctly distinguished based on landmark features.

Quantitative evaluation indicates strong performance in terms of accuracy, precision, recall, and F1-score. The model achieves reliable classification results across different gesture classes. As expected in real-time systems, slight variations in hand positioning may affect predictions, but overall performance remains consistent due to normalized feature representation.

The system demonstrates robustness in handling variations in illumination, orientation, and background conditions. The use of normalized landmark coordinates reduces dependency on image-specific features, improving generalization capability.

Real-time testing through the web-based interface confirms that the system can effectively convert hand gestures into text, allowing users to form words and sentences interactively, thereby enhancing the usability.

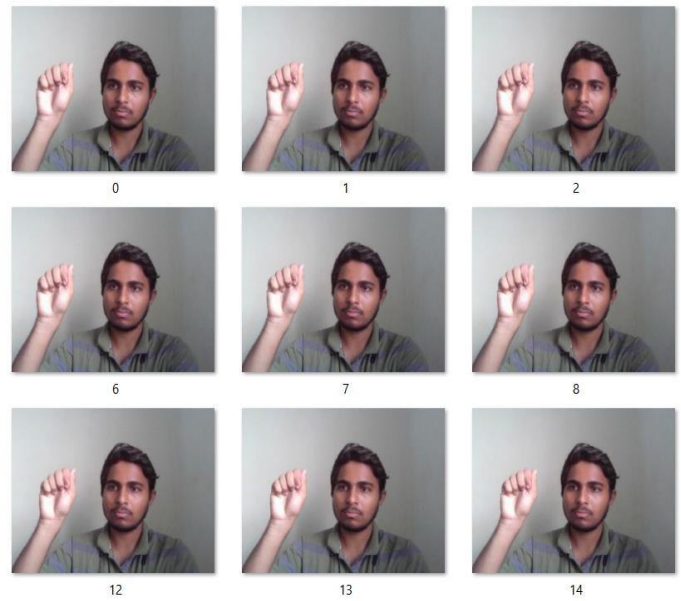


Figure - 3 : Datasets of letter - A



Figure - 4 : Datasets of letter - L

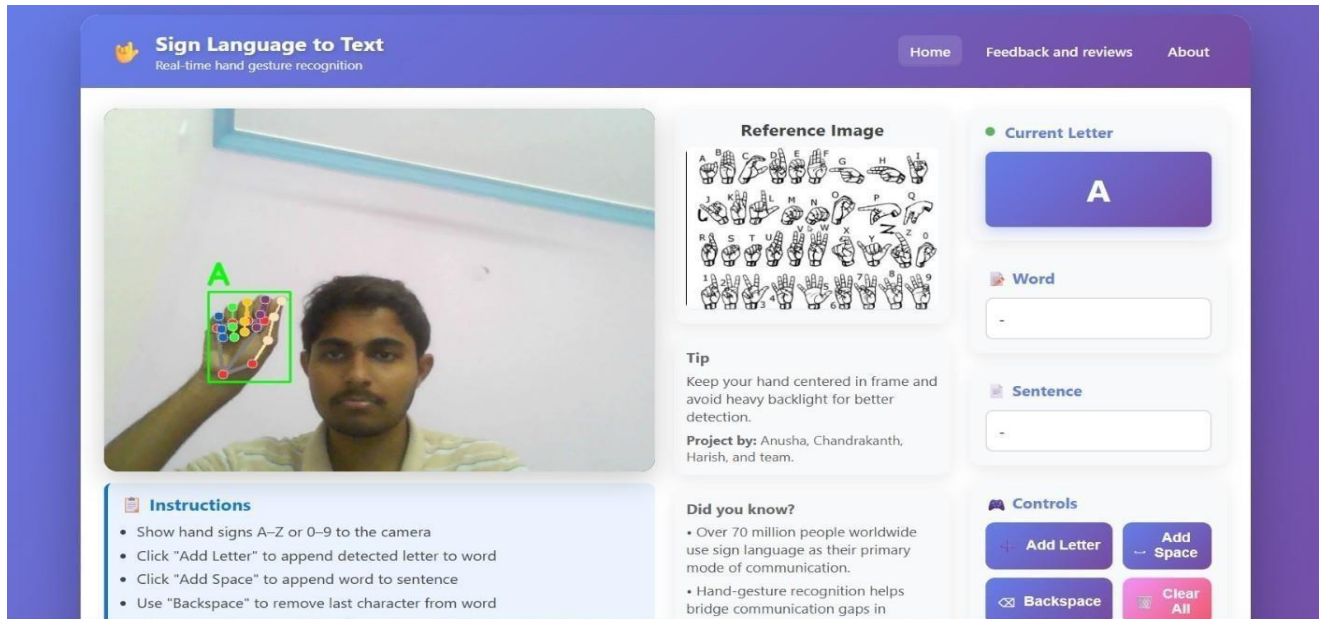


Figure - 5 : Real-Time Recognition of Sign Language Letter "A"

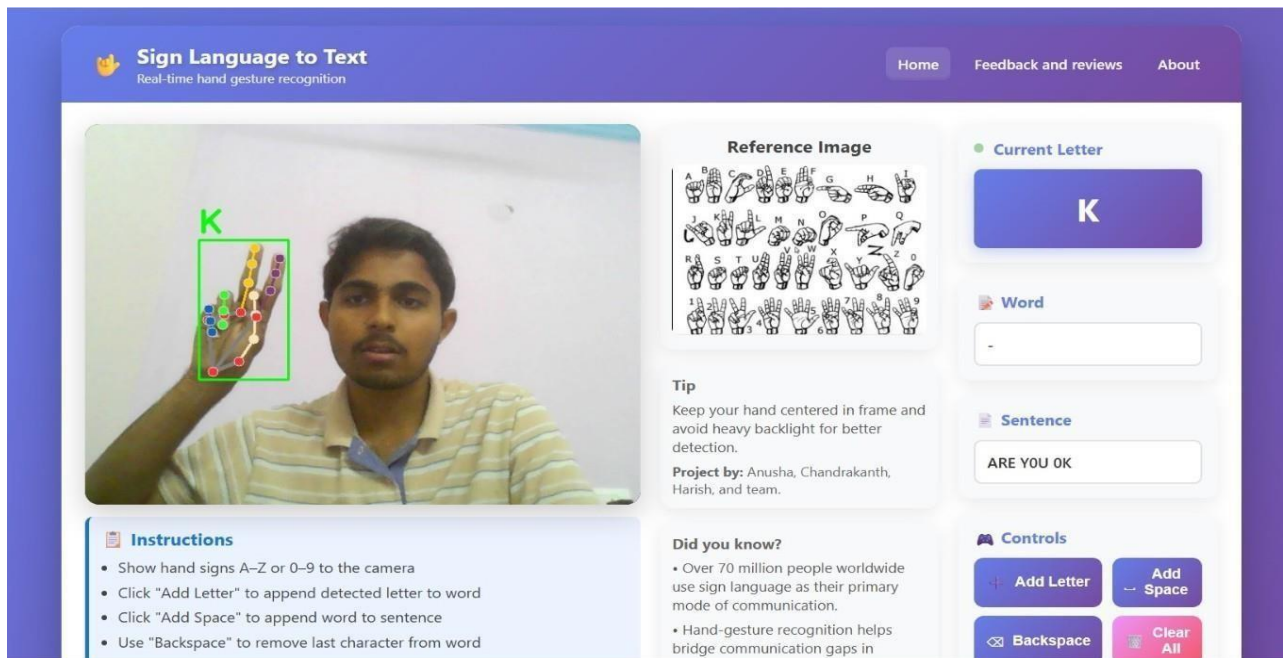


Figure - 6 : Real-Time Recognition of Sign Language Letter "K"

Table 1 - Execution Time Table

S.No	Metric	Value(%)
1	Accuracy	96.8
2	Precision	96.5
3	Recall	96.2
4	F1-Score	96.3

Table 2- Performance Table

Process	Time
Model Loading	3.8 sec
Hand Detection	1 sec
Gesture Prediction	4 sec
Result Rendering	2.4 sec
Total Execution Time	11.2 sec

7. Discussion

The results suggest that the proposed sign language recognition system is well-suited for real-time gesture-to-text conversion under practical conditions.

However, certain limitations were observed. Variations in hand positioning, lighting conditions, and background complexity can occasionally affect prediction accuracy. Gestures with similar finger configurations may sometimes lead to minor misclassifications.

Additionally, since the evaluation relies on classification-based metrics rather than controlled benchmark datasets, the reported performance reflects practical system behavior rather than ideal conditions. Future studies may incorporate larger and more diverse datasets evaluation.

8. Conclusion

This study presented a practical sign language recognition system based on MediaPipe and a Random Forest classifier, aimed at enabling real-time conversion of hand gestures into text. The primary objective of the work was to develop an efficient and reliable system capable of recognizing hand signs representing alphabets (A-Z) and digits (0-9) under real-world conditions. To achieve this, the proposed system incorporates a structured data collection process followed by feature extraction using normalized hand landmarks and classification through a machine learning model. The system is further integrated into a web-based interface to support real-time interaction and usability.

The results obtained from the experiments demonstrate that the proposed gesture recognition system achieves reliable performance in terms of accuracy and real-time responsiveness. The system effectively detects hand landmarks and classifies gestures with minimal computational complexity. The ability to convert gestures into words and sentences enhances its practical applicability in assistive communication. These improvements are particularly valuable in real-world scenarios where seamless and efficient interaction is required. Both qualitative observations and quantitative evaluation indicate that the proposed system is capable of providing accurate and consistent gesture recognition under varying conditions.

Overall, the proposed system demonstrates that real-time hand gesture recognition using MediaPipe and machine learning techniques can play an important role in improving assistive communication systems. By converting hand gestures into meaningful text, the framework contributes to enhancing accessibility and interaction for hearing and speech-impaired individuals, supporting the broader field of human-computer interaction and gesture-based communication.

Future research may focus on several possible improvements to further enhance the system's performance and applicability. One potential direction is the development of deep learning-based models such as Convolutional Neural Networks to improve recognition accuracy for complex gestures. Another important extension would involve expanding the dataset with more diverse hand gestures, varying backgrounds, and different lighting conditions to improve generalization. Additionally, integrating the system with real-time video processing pipelines.

References

- 1) Zhang, Y., & Jiang, X. (2024). "Recent Advances on Deep Learning for Sign Language Recognition." *Computer Modeling in Engineering & Sciences*.
- 2) Ansar, H., Al Mudawi, N., Alotaibi, S. S., Alazeb, A., Alabdullah, B. I., Alonazi, M., & Park, J. (2023). "Hand Gesture Recognition for Characters Understanding Using Convex Hull Landmarks and Geometric Features." *IEEE Access*.
- 3) Rajalakshmi, E., Elakkiya, R., Subramaniaswamy, V., Alexey, P., Mikhail, G., Bakaev, M., Kotecha, K., Gabralla, L. A., & Abraham, A. (2023). "Multi-Semantic Discriminative Feature Learning for Sign Gesture Recognition Using Hybrid Deep Neural Architecture." *IEEE Access*.
- 4) Rokade, Y. I., & Jadav, P. M. (2017). "Indian Sign Language Recognition System." *International Journal of Engineering and Technology*.
- 5) Reshna, S., & Vidhya, K. V. (2023). "Recognition of Indian Sign Language using Hand Gestures and Facial Expressions." *International Conference Paper*.
- 6) Sneha, B. S., Sowmya, R., Srilakshmi, T. M., Bhat, S., & Reddy, S. (2022). "Sign Language Recognition System Using Indian Sign Language." *International Journal of Creative Research Thoughts (IJCRT)*.
- 7) Shinde, A., & Kagalkar, R. (2014). "Sign Language Recognition for Deaf Sign User." *International Journal for Research in Applied Science & Engineering Technology*.
- 8) Goyal, S., Sharma, I., & Sharma, S. (2013). "Sign Language Recognition System For Deaf And Dumb People." *International Journal of Engineering Research & Technology*.
- 9) Singha, J., & Das, K. (2015). "Automatic Indian Sign Language Recognition for Continuous Video Sequence." *ADBU Journal of Engineering Technology*.
- 10) Rokade, Y. I., & Jadav, P. M. (2017). "Indian Sign Language Recognition System using Vision-Based Approach." *International Journal of Engineering and Technology*.