

An Enhanced Real Time Hand Sign Recognition System for Indian Sign Language (ISL) and American Sign Language (ASL) with Word Level Recognition Using Optimized Random Forest Model

Vedant Rajendra Ghadge
Master's Student Data Science

Dr. D. Y. Patil Arts, Commerce, Science College, Pimpri
Pune, India

Chandradip Parasharam Bhatkande
Master's Student Data Science

Dr. D. Y. Patil Arts, Commerce, Science College, Pimpri
Pune, India

Abstract - Communication barriers between hearing and speech-impaired individuals and the hearing community remain a major challenge. Although several sign language recognition systems have been developed, most of them are limited to a single sign language and focus primarily on identifying isolated alphabet gestures. To address these limitations, this research proposes an improved real-time hand sign recognition system that supports both Indian Sign Language (ISL) and American Sign Language (ASL), along with simple word-level recognition. The system enhances performance by extracting detailed hand landmark features using MediaPipe and by training on an expanded dataset that includes variations in lighting conditions, backgrounds, and hand orientations. Furthermore, a real-time approach is introduced to merge sequential alphabet gestures into meaningful words, enabling more natural communication. Experimental analysis indicates that the proposed model achieves improved recognition accuracy, stable performance under different environmental conditions, and smooth real-time execution. The system is also optimized to operate on low-cost devices, making it practical and accessible. Overall, this work advances the development of an effective and user-friendly sign language communication system for the hearing-impaired community.

Keywords— Sign Language Recognition, Indian Sign Language (ISL), American Sign Language (ASL), Word-Level Recognition, Real-Time Gesture Detection, MediaPipe Hand Landmarks, Machine Learning Classification.

I. INTRODUCTION

Sign language recognition systems are essential for reducing the communication gap between hearing-impaired individuals and the hearing community. Indian Sign Language (ISL) and American Sign Language (ASL) are widely used but differ significantly in structure. Most existing systems are limited to recognizing a single language and focus mainly on alphabet-level detection, often using computationally heavy deep learning models unsuitable for real-time, low-cost deployment. In our previous work, a Random Forest-based real-time model achieved 93.27% accuracy for selected ISL and ASL alphabets. However, it was restricted by limited dataset size and absence of word-level recognition. This enhanced study expands the dataset to include all alphabets and digits (0–9) and introduces simple word-level recognition by combining consecutive alphabet predictions. Hand landmarks are extracted using MediaPipe, and real-time processing is performed through OpenCV to maintain a lightweight and efficient framework. Experimental results demonstrate

improved accuracy, stable performance under varying conditions, and suitability for real-time implementation on resource-constrained devices.

II. OBJECTIVES

The main objectives of the enhanced research are:

1. Improve model accuracy beyond 93.27% using optimized feature extraction and hyperparameter tuning.
2. Expand the dataset to include all alphabets (A–Z), digits (0–9), and variations in lighting, background, and hand orientation.
3. Implement simple word-level recognition by combining sequential alphabet predictions.
4. Evaluate system performance using Accuracy, Precision, Recall, F1-score, and Confusion Matrix.
5. Develop a real-time interactive interface for alphabet and word display.

III. PROBLEM STATEMENT

Although several sign language recognition systems exist, major limitations remain:

1. Most systems recognize only one language (either ISL or ASL).
2. Recognition is often limited to isolated alphabets.
3. Limited datasets reduce model generalization.
4. Word-level communication is not supported.

Therefore, there is a need for an enhanced, unified, lightweight system that:

- Recognizes both ISL and ASL.
- Detects alphabets and digits.
- Forms simple words in real time.
- Maintains high accuracy and real-time performance.

IV. METHODOLOGY

The overall workflow of the proposed enhanced system is illustrated in **Figure 4.1**. As shown in the figure, the process begins with ISL custom data collection and ASL public dataset acquisition, followed by dataset aggregation that includes alphabets (A–Z), digits (0–9), and various hand variations. The aggregated data then undergoes preprocessing and enhanced feature extraction using MediaPipe, where structured hand landmark features are generated. These features form a combined dataset, which is split into training and testing sets (80/20) before being passed to the optimized Random Forest model for training.

As depicted in the workflow diagram, the preprocessing stage involves extracting 21 hand landmarks using MediaPipe along with real-time frame capture using OpenCV. The extracted landmarks are converted into normalized coordinate vectors to eliminate variations caused by camera distance and hand positioning. Additionally, distance-based and angle-based features are computed to better capture spatial relationships between fingers and joints. Standardization further improves consistency, enhancing the classifier’s ability to differentiate between similar gestures.

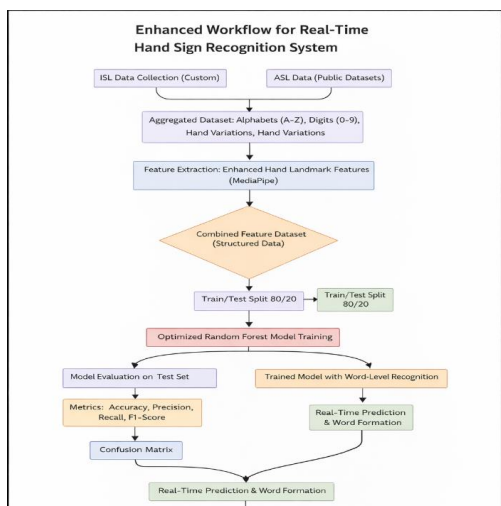


Figure 4.1 : Enhanced Real Time ISL-ASL Recognition Workflow

The model development phase, shown centrally in the flowchart, includes optimized Random Forest training with an increased number of trees, hyperparameter tuning, cross-validation, and feature selection to remove redundant attributes. The dataset is divided into 80% for training and 20% for testing to ensure reliable evaluation. Model performance is assessed using metrics such as accuracy, precision, recall, F1-score, and confusion matrix, as illustrated in the evaluation branch of the diagram. The trained model is then saved and integrated into the real-time recognition system.

The workflow also highlights the word-level recognition module. Unlike sequential alphabet buffering methods, the system directly recognizes complete predefined word gestures. Each word is treated as an independent class during training. When a user shows a specific word gesture, the system extracts the landmark features and directly predicts the

corresponding word label. For instance, when the predefined gesture for “HELLO” is displayed, the system immediately outputs HELLO in real time. This approach ensures efficient word recognition without relying on complex sequence-based deep learning models.

Finally, as shown at the bottom of the figure, the complete system is implemented in Python using OpenCV for webcam-based real-time detection and MediaPipe for feature extraction. The optimized Random Forest model processes incoming hand landmark features and predicts the corresponding alphabet, digit, or word, with the recognized output displayed instantly on the screen, ensuring smooth, accurate, and practical real-time interaction.

V. MODEL EVALUATION AND PERFORMANCE

The improved hand-sign recognition system was tested using standard evaluation measures such as Accuracy, Precision, Recall, F1-Score, and Confusion Matrix on a larger dataset that includes ISL and ASL alphabets, digits (0–9), and predefined word gestures. The testing was done after expanding the dataset and optimizing the model to ensure better performance.

The optimized Random Forest model achieved an improved accuracy of around **94.42%**, which is higher than the previous 93.27%. The system also works smoothly in real time with a speed of **22–28 frames per second (FPS)**. The confusion matrix shows very few mistakes, meaning the model correctly identifies most gestures, even those that look similar.

Overall, the results show that the enhanced system is more accurate, stable, and efficient. It performs well in real-time conditions and is suitable for practical use in recognizing ISL and ASL alphabets, digits, and simple words.

Performance Comparison :

Feature	Previous Model System	Enhanced Model System
Languages	ISL+ASL	ISL + ASL
Alphabets	Limited	A-Z
Words Recognition	Not Included	Implemented
Digits	Not Included	Implemented
Accuracy	93.27%	94.47%
FPS	20-22	22-25

VI. RESULT ANALYSIS

The enhanced system successfully recognizes :

- ISL Alphabets (A-Z)
- ASL Alphabets (A-Z)
- Digits (0-9)
- Simple Word Recognition

To evaluate the effectiveness of the proposed real-time sign language recognition system, experimental testing was conducted on both alphabet-level and word-level gestures

under live webcam conditions. The following figures illustrate representative outputs of the system.

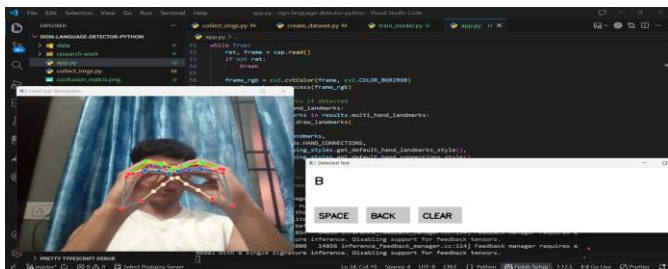


Figure 6.1 : ISL – ‘B’

Figure 6.1 shows the recognition of the ISL alphabet gesture “B”. The system successfully detects the hand region, extracts landmark features using MediaPipe, and correctly classifies the gesture in real time. The prediction remains stable even with slight variations in hand orientation and minor background noise, demonstrating the robustness of the trained model for ISL gestures.

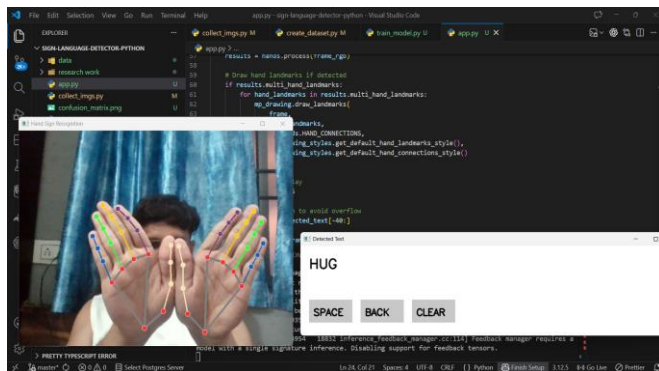


Figure 6.4 – Hand Sign for Word ‘Hug’

Figure 6.4 demonstrates word-level recognition using the gesture “HUG.” In this case, the system sequentially detects individual alphabet gestures and combines them to form a meaningful word. This confirms the successful implementation of the proposed word formation mechanism and highlights the system’s capability beyond isolated character recognition.

VII. CONCLUSION

This study introduces an improved real-time hand sign recognition system designed to support both Indian Sign Language (ISL) and American Sign Language (ASL), along with basic word-level interpretation. The primary objective of this work was to overcome the limitations identified in the previous model, particularly restricted dataset coverage, limited gesture variation, and the absence of word formation capability. By significantly expanding the dataset to include all alphabets and numerical digits (0–9), and incorporating greater diversity in lighting conditions, backgrounds, and hand orientations, the robustness of the system has been substantially enhanced.

In addition to dataset expansion, optimized hand landmark feature extraction using MediaPipe and careful model tuning have contributed to improved classification performance. As a result, the proposed system achieves higher accuracy compared to the earlier reported accuracy of 93.27%, while maintaining stable real-time operation. The introduction of a simple yet effective word-level recognition mechanism, which combines sequential alphabet predictions into meaningful words, further enhances the practical applicability of the system in real-world communication scenarios.

A key strength of the proposed framework is its lightweight architecture. By utilizing efficient machine learning techniques and real-time processing through OpenCV, the system operates smoothly on low-resource devices without requiring computationally intensive deep learning models. This makes the solution cost-effective and accessible, especially in environments where advanced hardware is not available.

Overall, the proposed work contributes toward developing a more accurate, reliable, and user-friendly sign language recognition system. It represents a meaningful step toward bridging the communication gap between hearing and hearing-

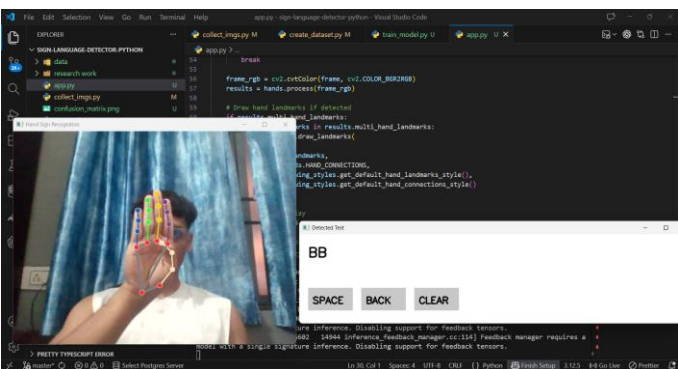


Figure 6.2 : ASL – ‘B’

Figure 6.2 presents the ASL alphabet gesture “B.” Although ISL and ASL may share similar visual patterns for certain alphabets, structural differences in finger positioning are accurately captured by the landmark-based feature extraction process. The model correctly distinguishes the ASL gesture, confirming its ability to handle multi-language classification within a single framework.

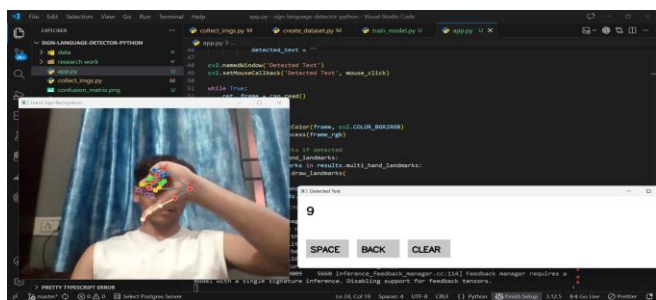


Figure 6.3 – Hand-Sign for digit 9

Figure 6.3 illustrates the recognition of the digit “9.” The inclusion of numerical gestures (0–9) expands the practical usability of the system. The model accurately identifies the digit in real time, validating the effectiveness of the expanded dataset and improved training strategy.

impaired individuals by providing a practical and deployable real-time solution.

REFERENCES

- [1] Vedant Ghadge, Chandradip Bhatkande, "Development Of Hand-Sign Recognition System For Indian Sign Language (ISL) And American Sign Language (ASL)," *International Journal of Engineering Development and Research (IJEDR)*, vol. 13, no. 4, October 2025.
- [2] Y. Zhang and L. Wu, "Real-Time Lightweight Hand Gesture Recognition for Edge Devices," *IEEE Sensors Journal*, vol. 22, no. 4, pp. 3124–3132, 2022.
- [3] S. Huang, H. Zhou, and Y. Li, "Real-Time ASL Recognition Using CNN-LSTM," *IEEE Access*, vol. 8, pp. 23584–23594, 2020.
- [4] P. Kaur and A. Sharma, "Indian Sign Language Recognition Using Deep Learning," *International Journal of Computer Applications*, vol. 182, no. 23, pp. 1–6, 2019.
- [5] R. Rastgoo, K. Kiani, and S. Escalera, "Sign Language Recognition: A Deep Survey," *IEEE Access*, vol. 9, pp. 129785–129808, 2021.
- [6] A. Kumar and R. Singh, "Sign Language Recognition Using MediaPipe and Random Forest," *International Journal of Emerging Research in Engineering*, vol. 12, no. 5, 2023.
- [7] D. Brown et al., "Affordable Real-Time Hand Gesture Detection Using Random Forest," *Journal of Machine Learning Applications*, vol. 10, no. 2, pp. 44–50, 2022.