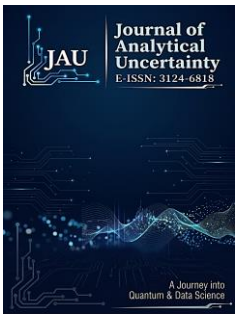
OPEN ACCESS PEER-REVIEWED

JAU Vol.2, No.1, 2026 - Page: 121-132
doi.org/10.63924/jau.v2i1.328
e-ISSN: 3124-6818



Check for updates

Real Time Indian Sign Language Recognition System Using Mediapipe and Multi-Layer Perceptron

Binod Kumar Sharma[1*], Adarsh Sagar[2], Manoj Kumar Das[3], Mayank Jaiwal[4], Aditya Kumar[5], Abhijeet Kumar[6]

[1]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India; [2]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India; [3]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India; [4]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India; [5]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India; [6]Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India

*Corresponding author: Binod Kumar Sharma, Department of Computer Science, Birla Institute of Technology, Mesra, Ranchi, India. Email: bksharma@bitmesra.ac.in

Abstract

Indian Sign Language (ISL) is an essential means of communication for individuals with hearing and speech impairments. However, communication barriers still exist due to limited availability of real-time sign language recognition systems and lack of public awareness. To address these challenges, this paper presents a real time Indian Sign Language recognition system that combines computer vision and machine learning techniques to recognize ISL alphabets and numerical gestures. The proposed system utilizes MediaPipe Hands to detect and extract three-dimensional hand landmarks coordinates from live webcam input, which are then classified using a multi-layer perceptron (MLP) model. A custom dataset of ISL gestures was created, preprocessed, and used for model training and evaluation. The trained model is integrated into a web-based application that enables users to perform gesture recognition in real time without requiring any wearable devices or hardware. Experimental evaluation demonstrated that the proposed system provides efficient real time recognition under normal conditions while maintaining low computational complexity. The developed framework offers a practical, cost-effective, and user-friendly solution for supporting ISL learning and improving communication among deaf and hearing communities.

Keywords: Indian Sign Language, Sign Language Recognition, MediaPipe, Computer Vision, Multi-Layer Perceptron, Machine Learning, Gesture Recognition

Introduction

Sign language is one of the most effective communication methods for individuals with hearing and speech impairments [1]. Unlike spoken languages, it relies on hand gestures, facial expressions, and body movements to convey information. However, sign languages are not universal; each country has its own sign language [2]. In India, Indian Sign Language (ISL) serves as the primary communication medium for the deaf community, but it lacks awareness and creates significant barriers between hearing and non-hearing individuals.

Recent developments in computer vision and machine learning have helped us to develop gesture recognition systems capable of translating sign language into textual information. These technologies have the potential to improve accessibility, education, and everyday communication for people with hearing impairments [3].

Copyright: © 2026 by the author(s). This is an open access article distributed under the CC BY-SA 4.0 license

Received: May 09, 2026 - Accepted: November 22, 2026 - Published online: November 24, 2026

In this work, we present a real time Indian Sign Language recognition system designed to recognize ISL alphabets and numerical gestures using MediaPipe hand landmark detection and a Multi-Layer Perceptron (MLP) classifier. This system captures live video through a webcam, extracts hand landmarks, and performs real time gesture classification.

The major contributions of this work are summarized as follows:

- Development of a real time ISL recognition system based on computer vision.
- Extraction of hand landmarks using the MediaPipe framework.
- Design and training of a Multi-Layer Perceptron classifier for recognizing ISL alphabets and numbers.
- Deployment of a browser-based application for real time gesture recognition.

Related Work

Sign language recognition (SLR) has attracted considerable attention in the field of computer vision, pattern recognition, and human-computer interaction. Developments in machine learning and deep learning help to improve the accuracy and robustness of gesture recognition systems.

Several surveys have analyzed the process of sign language recognition. R. Rastgoo [1] reviewed recent developments in gesture segmentation and classification techniques, highlighting the transition from conventional image processing methods to deep learning-based approaches. Similarly, A. Wadhawan [2] discussed the application of computer vision techniques for recognizing hand gestures under different environments. A. Adeyanju [3] presented a broad view covering research conducted between 2001 and 2021, highlighting the growing adoption of artificial intelligence in sign language recognition.

Research focusing specifically on Indian Sign Language has explored different approaches for improving recognition performance. S. Ghotkar [4] investigated challenges associated with ISL recognition, including background complexity and skin tone. A. Tyagi [5] proposed vision-based ISL recognition methods using hand tracking and machine learning algorithms.

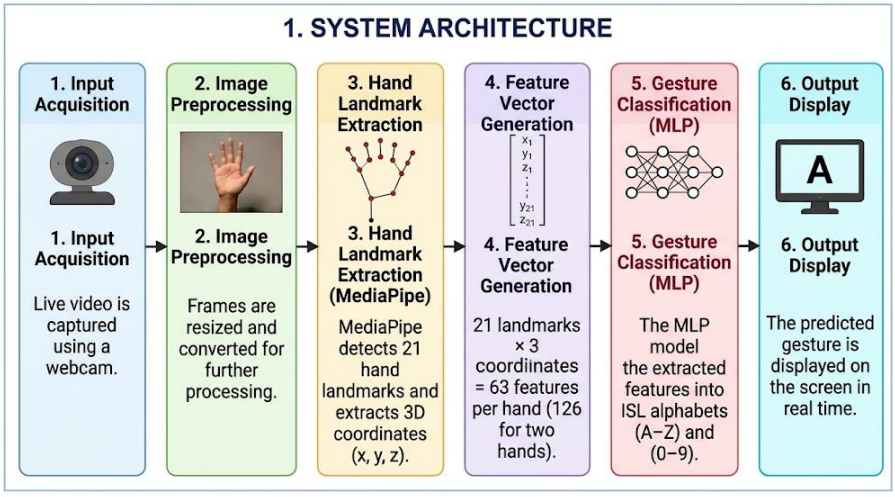
Traditional sign language recognition systems primarily relied on image preprocessing, color segmentation techniques such as RGB and HSV, and conventional machine learning classifiers. These methods achieved a certain level of performance under controlled conditions, though their accuracy decreased in real-world conditions. Recent developments have shifted toward landmark-based feature extraction and deep learning [6] models which improve accuracy.

By looking at recent developments, the proposed work develops a real time Indian Sign Language recognition system using MediaPipe [7] hand landmark extraction and a Multi-Layer Perceptron (MLP) system.

Proposed Methodology

The proposed system is a real time Indian Sign Language (ISL) recognition system (Figure 1). The system integrates computer vision [8] and machine learning techniques to achieve accurate and efficient gesture recognition without requiring any sensors.

Figure 1: System Architecture



The complete workflow consists of five major stages: dataset preparation, image preprocessing, hand landmark extraction, gesture classification, and real time prediction.

Dataset preparation

A custom dataset was developed to train the model. The dataset consists of images representing Indian Sign Language alphabets (A–Z) and numerical gestures (0–9). Images were collected from multiple sources to increase diversity in hand shapes, lighting conditions, and backgrounds.

Image preprocessing

Before training the classification model, all collected images were preprocessed to improve consistency and accuracy. The preprocessing stages included image resizing, color conversion, normalization, and removal of irrelevant background.

Feature extraction using mediapipe

MediaPipe detects the hand region and extracts 21 landmarks from each hand. Since each landmark contains three-dimensional coordinates (X, Y, Z), this yields a total of 63 features for each hand. For two hands, the extracted features were combined to form the final feature vector. Each detected hand landmark is represented as a three-dimensional coordinate:

$$L_i = (x_i, y_i, z_i), \quad i = 1, 2, \ldots, 21 \qquad (1)$$

where x_i, y_i, and z_i denote the normalized coordinates of the i-th hand landmark detected by the MediaPipe Hands framework (Figure 2 - 3). The feature vector for one hand is expressed as:

Figure 2: Creating Dataset

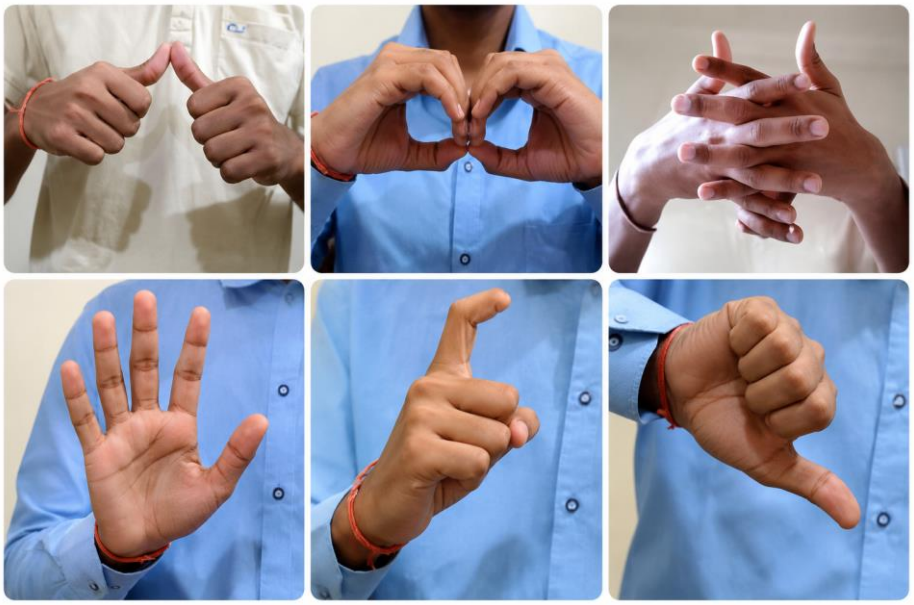
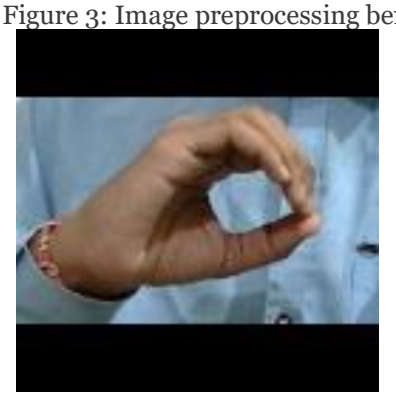
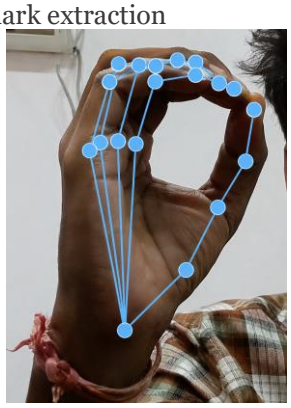


Figure 3: Image preprocessing before hand landmark extraction



(a) Original hand image



(b) Preprocessed Image

(2)

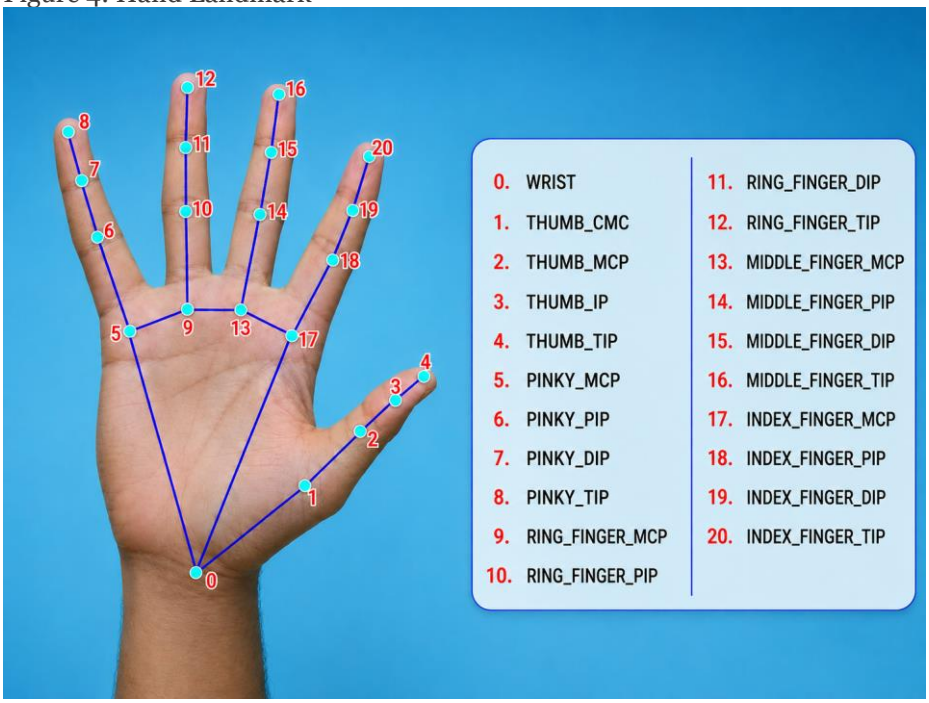
$F = [x_1, y_1, z_1, ..., x_{21}, y_{21}, z_{21}]$

Since we use both hands, the complete feature vector for two hands becomes:

$F_total = [F_left, F_right]$ (3)

resulting in a total of 126 input features (Figure 4).

Figure 4: Hand Landmark



Gesture classification using multi-layer perceptron

The extracted landmark features are classified using a Multi-Layer Perceptron (MLP) neural network. The MLP consists of multiple fully connected hidden layers with Rectified Linear Unit (ReLU) [9] activation functions and dropout layers [10] to minimize overfitting. The dataset was divided into training and testing subsets before model development. The network learns the relationship between landmark coordinates and their corresponding ISL gestures (Figure 5).

The output of each hidden layer is computed as:

$h = \sigma(Wx + b)$ (4)

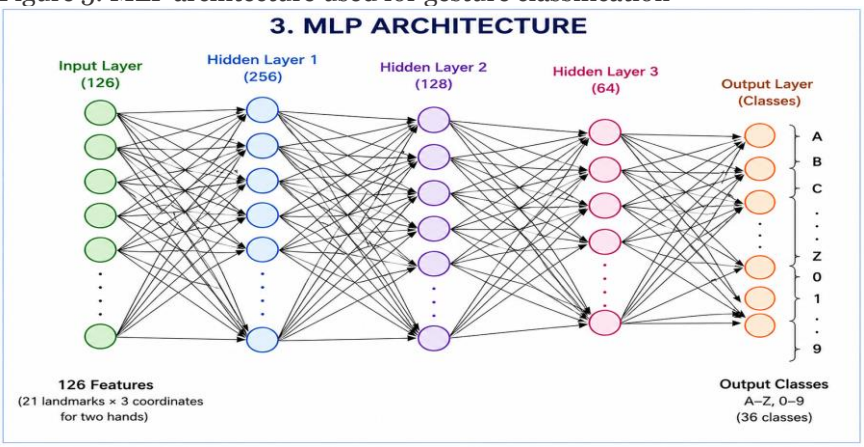
The Rectified Linear Unit (ReLU) activation function introduces non-linearity and is defined as:

$ReLU(x) = max(0, x)$ (5)

Dropout randomly deactivates neurons during training:

$y = 0$ *(with probability p),* $y = x / (1 - p)$ *(otherwise)* (6)

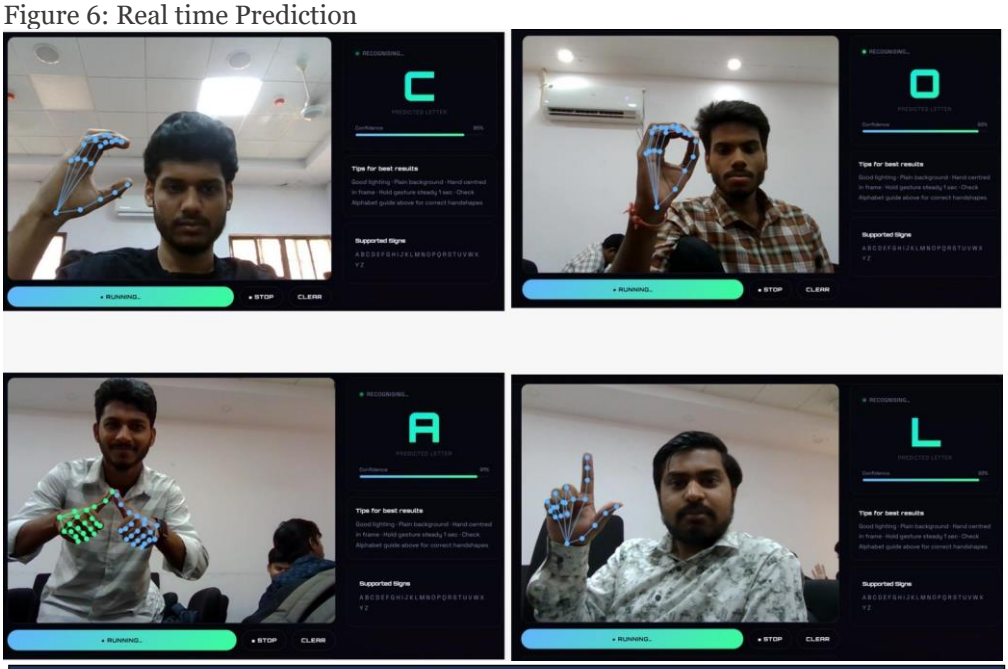
Figure 5: MLP architecture used for gesture classification



Real time gesture recognition

During system execution, live video is continuously captured using a webcam. Each frame is processed by MediaPipe to detect hand landmarks, which are transformed into numerical feature vectors identical to those used during training. The trained MLP classifier predicts the corresponding gesture label from the extracted landmarks. If the confidence of prediction satisfies the predefined criteria, it displays the alphabet or number to the user; otherwise, it classifies it as unknown (Figure 6).

Figure 6: Real time Prediction



Algorithm 1: Real-Time Indian Sign Language Recognition

```
Initialize webcam
Load the trained MLP model and label encoder
while webcam is active do
    Capture a video frame
    Detect hand landmarks using MediaPipe Hands
    if hand landmarks are detected then
        Extract 126 landmark features
        Normalize the feature vector
        Predict the gesture using the trained MLP model
        if prediction confidence exceeds the threshold then
            Display the predicted alphabet or number
        else
            Display "Unknown Gesture"
        end if
    end if
end while
Release webcam resources
```

Experimental Setup

The proposed Indian Sign Language (ISL) recognition system was implemented using Python. The experimental setup was designed to support real-time gesture recognition while maintaining low computational complexity. The implementation consists of four steps: dataset preprocessing, model training, model testing, and real-time deployment.

Software environment

The proposed system was developed using Python due to its extensive support for computer vision and machine learning libraries. OpenCV was used for image processing and preprocessing, while MediaPipe Hands was employed for real-time hand landmark detection (Table 1).

Table 1: Software Environment

Software	Version
Python	3.10
OpenCV	4.11

MediaPipe	Latest
PyTorch	2.x

Hardware configuration

The experiments were conducted on a standard laptop equipped with an Intel Core i3 (or higher) processor, 8 GB RAM, a 256 GB SSD, and an integrated webcam (Table 2). No dedicated GPU or specialized hardware was required, demonstrating that the proposed system can operate efficiently on commonly available computing devices.

Table 2: Hardware Requirements

Component	Specification
CPU	Intel Core i3 / i5
RAM	8 GB
Storage	256 GB SSD
Camera	Integrated Webcam

Dataset preparation

A custom dataset containing Indian Sign Language alphabets and numerical gestures was collected from verified sources. The dataset was organized into class-wise directories, with each directory representing a unique gesture category (Table 3). During preprocessing, MediaPipe Hands extracted landmark coordinates from every image, generating numerical feature vectors that served as input for the classifier. Invalid samples in which hand landmarks could not be detected were discarded to improve data quality.

Table 3: Dataset Statistics

Parameter	Value
Number of Classes	36 (A–Z + 0–9)
Total Images	81,000
Images per Class	2,400
Image Resolution	128 × 128
Dataset Source	Custom Captured
Landmark Features	126
Training Split	80%
Testing Split	20%

Table 3 summarizes the characteristics of the dataset used for training and evaluation. The dataset contains 81,000 custom captured images distributed across 36 gesture classes, with 2,400 images per class. Each image was resized to 128 × 128, and 126 landmark features were extracted (63 features per hand). The dataset was divided into training and testing data in an 80:20 ratio.

Model training

The extracted landmark features were divided into training and testing datasets before model development. The gesture labels were encoded into numerical form using a label encoder. The MLP classifier was trained using supervised learning with the Adam optimizer [11] and Cross-Entropy Loss function. Dropout layers were incorporated within the network architecture to reduce overfitting and improve generalization.

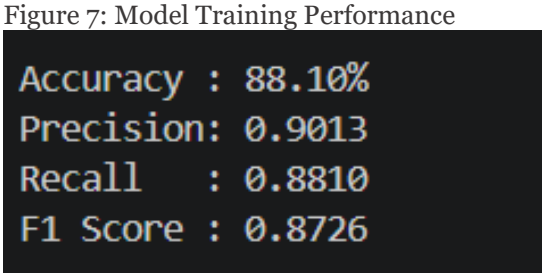
After training, the learned model parameters and label encoder were stored for future inference, allowing the trained classifier to be directly integrated into the real-time recognition system.

Table 4: MLP Model Configuration

Parameter	Value
Input Features	126
Hidden Layers	256, 128, 64
Activation Function	ReLU
Optimizer	Adam
Loss Function	Cross-Entropy Loss
Dropout Rate	0.3

Table 4 summarizes the configuration of the proposed Multi-Layer Perceptron (MLP) model. The network accepts 126 input features corresponding to the extracted hand landmark coordinates. It consists of three hidden layers containing 256, 128, and 64 neurons, respectively. The ReLU activation function is used to introduce non-linearity, while a dropout rate of 0.3 is used to reduce overfitting. The model is trained using the Adam optimizer with the Cross-Entropy Loss function for multi-class gesture classification (Figure 7).

Figure 7: Model Training Performance



Real-time testing

The trained model was evaluated using live webcam input. During execution, MediaPipe continuously detected hand landmarks from each captured video frame and generated feature vectors identical to those used during training. These features were passed to the trained MLP classifier, which predicted the corresponding ISL alphabet or numerical gesture in real time. The recognized gesture was immediately displayed through the web interface, enabling smooth interaction between the user and the recognition system. When the detected gesture did not sufficiently match any trained class, the system classified it as an unknown gesture, thereby reducing incorrect predictions. The experimental setup demonstrates that the proposed framework successfully integrates computer vision and machine learning techniques to provide efficient real-time Indian Sign Language recognition while maintaining low hardware requirements.

Results and Discussion

Experimental results

The proposed system was evaluated using real-time webcam input to assess its ability to recognize Indian Sign Language (ISL) alphabets and numerical gestures. During testing, the webcam continuously captured hand gestures, while MediaPipe Hands detected hand landmarks from each video frame. The extracted landmark coordinates were then provided to the trained Multi-Layer Perceptron (MLP) classifier for gesture prediction (Figure 8).

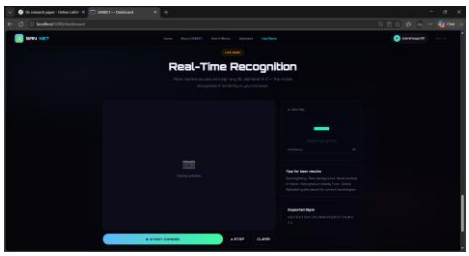
Experimental observations indicate that the proposed system successfully recognized trained ISL gestures under normal indoor lighting conditions. The prediction results were displayed instantly through the web interface, demonstrating the capability of the system to perform real-time gesture recognition with minimal delay. The use of landmark-based feature extraction significantly reduced the influence of background clutter, enabling stable recognition without requiring specialized sensors or wearable devices.

The developed web interface provides a simple and interactive environment where users can perform ISL gestures and immediately observe the corresponding predicted output. The real-time feedback makes the system suitable for educational and assistive communication applications.

Figure 8: Web interface of the proposed system



(a) About



(b) Live Demo

Performance analysis

The combination of MediaPipe hand landmark extraction and the Multi-Layer Perceptron classifier proved effective for recognizing static Indian Sign Language gestures. Instead of processing complete images, the proposed system relies on landmark coordinates, which considerably reduce computational complexity while preserving geometric information about hand posture (Figure 9-8).

The lightweight architecture enables real-time execution on standard computing hardware without the need for dedicated graphics processing units. This makes the proposed system computationally efficient and suitable for deployment on commonly available laptops and desktop computers.

Table 5: Overall Performance of the Proposed MLP Model

Metric	Value
Accuracy	88.10%
Precision	90.13%
Recall	88.10%
F1-Score	87.26%

Table 5 summarizes the overall evaluation metrics of the proposed ISL recognition model. The model achieved an accuracy of 88.10%, with a precision of 90.13%, recall of 88.10%, and F1-Score of 87.26%.

Figure 9: Confusion matrix

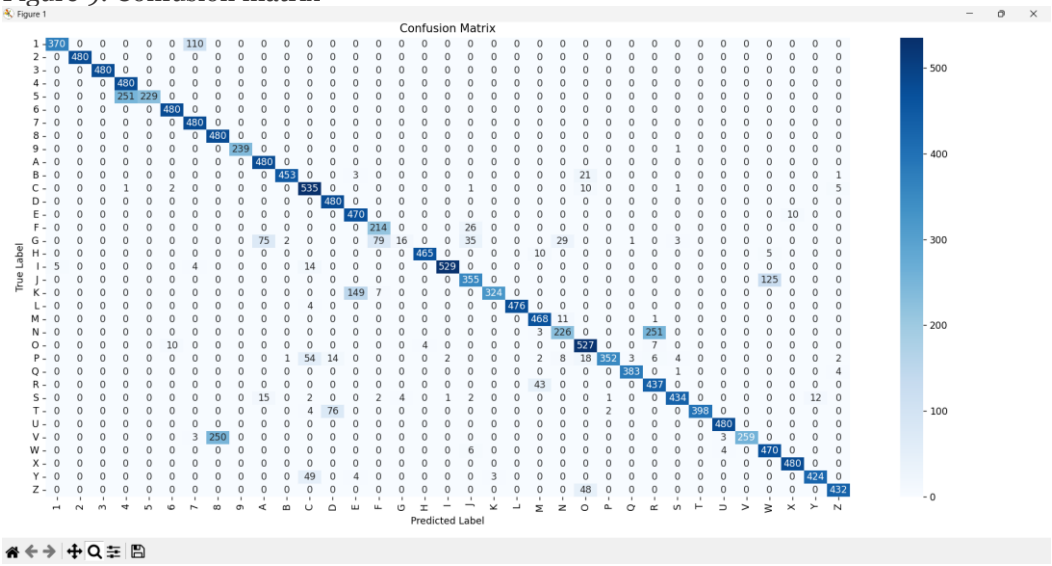


Figure 10: Training loss over 20 epochs

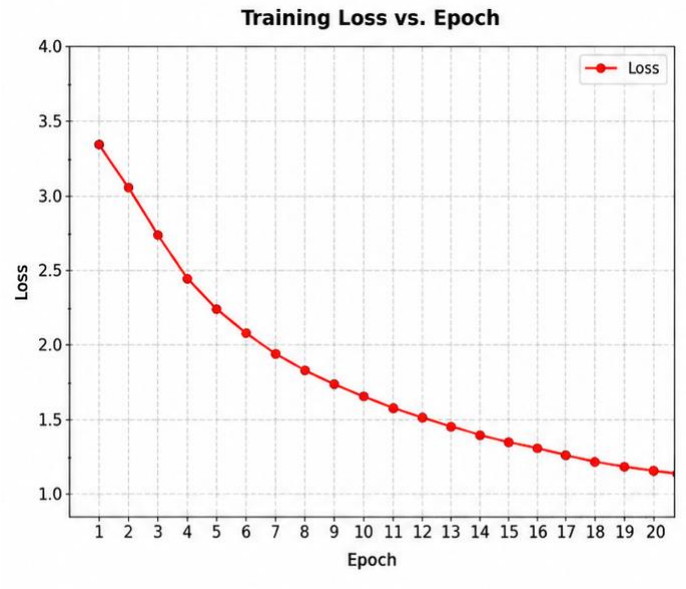
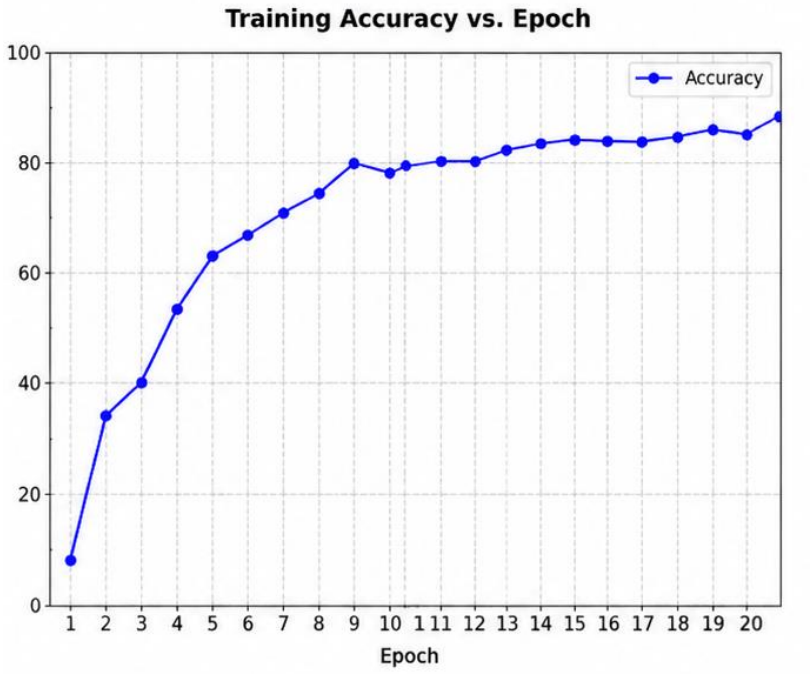


Figure 11: Training accuracy over 20 epochs



Discussion

Although the proposed system demonstrates reliable performance for recognizing static ISL alphabets and numerical gestures, several practical limitations were observed during experimentation (Table 6).

Recognition accuracy is highly dependent on image quality, illumination conditions, and proper visibility of the hand. Poor lighting, motion blur, partial hand occlusion, or incorrect hand positioning may reduce prediction accuracy. Additionally, since the current model has been trained primarily on static gestures, it cannot accurately recognize dynamic gestures involving continuous hand movement.

The system also relies on successful landmark detection by MediaPipe. If hand landmarks are not detected because of extreme viewing angles or incomplete hand visibility, gesture recognition cannot be performed correctly.

Despite these limitations, the proposed framework demonstrates that landmark-based feature extraction combined with a lightweight neural network provides an efficient solution for real-time Indian Sign Language recognition. Compared with conventional image-based approaches, the proposed system requires lower computational resources while maintaining practical usability for educational and assistive applications.

Table 6: Class-wise performance of the Model

Class	Precision	Recall	F1	Class	Precision	Recall	F1
1	0.9867	0.7708	0.8655	J	0.8353	0.7396	0.7845
2	1.0000	1.0000	1.0000	K	0.9908	0.6750	0.8030
3	1.0000	1.0000	1.0000	L	1.0000	0.9917	0.9958
4	0.6557	1.0000	0.7921	M	0.8897	0.9750	0.9304
5	1.0000	0.4771	0.6460	N	0.8248	0.4708	0.5995
6	0.9756	1.0000	0.9877	O	0.8446	0.9617	0.8993
7	0.8040	1.0000	0.8914	P	0.9915	0.7554	0.8575
8	0.6575	1.0000	0.7934	Q	0.9897	0.9871	0.9884
9	1.0000	0.9958	0.9979	R	0.6225	0.9104	0.7394
A	0.8421	1.0000	0.9143	S	0.9775	0.9175	0.9466
B	0.9934	0.9477	0.9700	T	1.0000	0.8292	0.9066
C	0.8082	0.9640	0.8792	U	0.9856	1.0000	0.9928
D	0.8421	1.0000	0.9143	V	1.0000	0.5029	0.6693
E	0.7508	0.9792	0.8499	W	0.7833	0.9792	0.8704

F	0.7086	0.8917	0.7897	X	0.9796	1.0000	0.9897
G	0.8000	0.0667	0.1231	Y	0.9725	0.8833	0.9258
H	0.9915	0.9688	0.9800	Z	0.9730	0.9000	0.9351
I	0.9944	0.9583	0.9760	—	—	—	—

Summary of results

The experimental evaluation demonstrates that the proposed system:

- Successfully performs real-time recognition of trained Indian Sign Language alphabets and numerical gestures.
- Accurately extracts hand landmarks using the MediaPipe Hands framework.
- Provides fast gesture prediction through a Multi-Layer Perceptron classifier.
- Operates efficiently on standard consumer hardware without requiring specialized equipment.
- Offers an intuitive web-based interface suitable for learning and practicing Indian Sign Language.

Applications

The proposed system has several practical applications in the fields of accessibility, education, and human-computer interaction. By enabling real-time recognition of Indian Sign Language (ISL) gestures, the system provides an effective communication aid for individuals with hearing and speech impairments while promoting inclusive interaction with the general public.

One of the primary applications of the proposed system is in educational institutions, where students can use the web-based platform to learn and practice Indian Sign Language interactively. The immediate visual feedback provided by the system enables learners to verify the correctness of their gestures, thereby improving the learning experience.

The proposed framework can also serve as an assistive communication tool in public places such as hospitals, banks, government offices, railway stations, airports, and customer service centers. In such environments, the system can help bridge the communication gap between hearing-impaired individuals and service providers without requiring the presence of a human interpreter.

Another important application is its integration into assistive technologies and smart communication platforms. Since the system is based on computer vision and operates using a standard webcam, it can be incorporated into desktop applications, web platforms, and other digital communication systems to facilitate real-time gesture interpretation.

Furthermore, the lightweight architecture of the proposed system makes it suitable for deployment on low-cost computing devices, thereby increasing its accessibility in educational and community environments where specialized hardware may not be available.

Although the current implementation focuses on recognizing static ISL alphabets and numerical gestures, the proposed framework establishes a strong foundation for future intelligent sign language translation systems capable of recognizing complete words and sentences.

Conclusion and Future Work

This paper presented a real-time Indian Sign Language recognition system based on MediaPipe hand landmark detection and a Multi-Layer Perceptron (MLP) classifier. The proposed framework combines computer vision and machine learning techniques to recognize alphabets and numerical gestures efficiently using a standard webcam without requiring wearable devices or specialized sensors.

The system successfully demonstrates that landmark-based feature extraction provides an effective representation of hand gestures while significantly reducing computational complexity compared to conventional image-based approaches. The developed web-based interface further improves usability by enabling users to perform real-time gesture recognition and receive immediate visual feedback.

Experimental evaluation indicates that the proposed system can accurately recognize trained ISL gestures under normal operating conditions while maintaining real-time performance on

standard computing hardware. The lightweight architecture and low hardware requirements make the system suitable for educational environments and assistive communication applications.

Despite its effectiveness, the current implementation has certain limitations. The system primarily recognizes static hand gestures and its performance may decrease under poor lighting conditions, hand occlusion, or rapid hand movement. In addition, recognition accuracy depends on successful hand landmark detection by the MediaPipe framework.

Future work will focus on extending the proposed framework to recognize dynamic Indian Sign Language gestures involving continuous hand movements and sentence-level interpretation. The incorporation of advanced deep learning architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformers, or attention-based models, may further improve recognition accuracy and robustness. Future enhancements may also include multilingual output generation, speech synthesis, mobile application deployment, and cloud-based services to support wider adoption of Indian Sign Language technology.

Overall, the proposed system represents a practical and cost-effective step toward improving accessibility and reducing communication barriers for individuals with hearing and speech impairments through intelligent real-time Indian Sign Language recognition.

Ethics approval

Not required.

Acknowledgments

No acknowledgements.

Competing interests

All the authors declare that there are no conflicts of interest.

Funding

This study received no external funding. Underlying data: Derived data supporting the findings of this study are available from the corresponding author on request.

Underlying data

Derived data supporting the findings of this study are available in this article.

CRediT authorship contribution statement

Binod Kumar Sharma, Adarsh Sagar, Manoj Kumar Das, Mayank Jaiwal, Aditya Kumar, Abhijeet Kumar: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Project administration, Supervision.

Declaration of artificial intelligence use

We hereby confirm that no artificial intelligence (AI) tools or methodologies were utilized at any stage of this study, including during data collection, analysis, visualization, or manuscript preparation. All work presented in this study was conducted manually by the authors without the assistance of AI-based tools or systems.

How to cite

Sharma, B. K., Sagar, A., Das, M. K., Jaiswal, M., Kumar, A., & Kumar, A. (2026). Real Time Indian Sign Language Recognition System Using Mediapipe and Multi-Layer Perceptron. *Journal of Analytical Uncertainty*, *2*(1), 121–132. https://doi.org/10.63924/jau.v2i1.328

References

[1] R. Rastgoo, B. Kiani, and S. Escalera, "Sign language recognition: A deep survey," Expert Systems with Applications, vol. 164, 2021.

[2] A. Wadhawan and P. Kumar, "Sign language recognition systems: A decade systematic literature review," Archives of Computational Methods in Engineering, 2021.
[3] I. A. Adeyanju et al., "Machine learning methods for sign language recognition: a critical review and analysis," Intelligent Systems with Applications, 2021.
[4] S. Ghotkar et al., "Hand gesture recognition for indian sign language," International Journal of Computer Applications, 2012.
[5] A. Tyagi et al., "Vision-based indian sign language recognition using machine learning," Procedia Computer Science, 2020.
[6] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
[7] C. Lugaresi, J. Tang, H. Nash et al., "Mediapipe: A framework for building perception pipelines," arXiv preprint arXiv:1906.08172, 2019.
[8] G. Bradski, "The opencv library," in Dr. Dobb's Journal of Software Tools, 2000.
[9] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," Proceedings of ICML, 2010.
[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," Journal of Machine Learning Research, vol. 15, pp. 1929–1958, 2014.
[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," International Conference on Learning Representations, 2015.