International Journal of Enhanced Research in Management & Computer Applications,
ISSN: 2319-7471, Volume 14 Issue 4, April 2025

Vision-Based Real-Time Recognition of Indian Sign Language Gestures Using Deep Learning Techniques

Anuj Rawat[1], Apurvi Ujjwal[2], Ashutosh Saraswat[3], Anusha Jain[4]

[1,2,3] Student, Department of Computer Science, Medi-Caps University, Indore – 453331, India
[4]Assistant Professor, Department of Computer Science, Medi-Caps University, Indore – 453331, India

ABSTRACT

Sign language provides a critical means of communication for individuals with hearing and speech impairments, distinguished by its unique visual-manual linguistic structure. However, communication barriers between signers and non-signers remain significant due to limited interpreter availability and high associated costs. This study presents a vision-based real-time Indian Sign Language (ISL) recognition system employing a Long Short-Term Memory (LSTM) model. A dataset of 36 static hand gestures, including 26 alphabetic (A–Z) and 10 numeric (0–9) signs, was compiled from multiple sources and augmented through webcam recordings. The system utilizes a conventional webcam for gesture capture, negating the need for specialized sensor devices and reducing overall costs. Achieving recognition accuracies between 70% and 80% under diverse lighting and background conditions, the system demonstrates robustness and practical applicability. This research advances Human-Computer Interaction by delivering an affordable, accessible tool to bridge communication gaps for the deaf and hard-of-hearing community. Future work will focus on integrating Natural Language Processing (NLP) techniques for dynamic gesture recognition and expanding the dataset for enhanced system performance.

Keywords: Sign Language Recognition, Human-Computer Interaction, Deep Learning, Indian Sign Language, Vision-Based Gesture Recognition, Long Short-Term Memory, Real-Time Processing, Accessibility Enhancement.

INTRODUCTION

Sign languages have been developed primarily to aid individuals who are deaf or mute in communicating effectively with others. They use a structured combination of hand gestures, hand shapes, and orientations to convey specific information. These visual-manual languages provide an essential means of communication for millions of people worldwide.

This project falls under the domain of Human-Computer Interaction (HCI) and aims to recognize multiple alphabets (A-Z), digits (0-9), and several common ISL (Indian Sign Language) hand gestures in real time. The recognition of hand gestures presents an advanced challenge particularly in ISL because gestures which need both hands raise recognition problems. The existing approaches which use glove sensors along with edge detection and the Hough Transform tend to be both cost-intensive and unavailable for widespread use in the population.

People who are deaf or hard-of-hearing form a significant percentage of Indian population and make ISL their primary communication tool. The low level of sign language understanding among the population establishes a communication difficulty that needs an interpreter's support at a high cost alongside significant inconvenience. This research develops AI software to analyze ISL hand signals in real time so people with hearing disabilities can communicate easily with the general population.

PROBLEM STATEMENT

The condition known as hearing impairment affects numerous people throughout the world making it a leading sensory disability. Academic research indicates approximately how many Indians experience hearing and speech challenges which present substantial obstacles for their communication abilities. The communication barrier intensifies because most people who do not have hearing challenges fail to learn sign language thus creating communication barriers for those who need it.

People who have hearing or speech disabilities encounter various communication problems because there is no affordable and effective technique to translate sign languages. Current sensor-based solutions together with image processing methods have shown development but remain too complex and expensive while being unattainable for regular people. The expensive price of commercial solutions prevents middle-class families as well as those who cannot afford high-cost products from utilizing them effectively.

This research develops an AI-based system for real-time sign language detection through computer vision and machine learning which seeks to close communication barriers. The combination of webcam functionality with deep learning models permits the system to read Indian Sign Language gestures that feed into neural networks before displaying the translated text on screen thus establishing communication between non-signers and signers.

This system differs from other speech-to-text solutions that help speakers who hear because it caters specifically towards users who cannot both speak and hear. The development aims to achieve accessible sign language recognition technology that also works without expensive hardware systems while remaining simple to use by all people and enabling free communication between hearing and speech impaired individuals without interpreter requirements..

LITERATURE REVIEW

The research field of hand gesture recognition has expanded significantly since recent decades because of its importance to computer vision and machine learning as well as human-computer interaction (HCI). The main target of gesture recognition systems involves changing detected hand movements into operational commands or written results. Scientists have designed multiple methods involving sensors and visual detection systems for recognizing sign language in real time.

Hand gesture recognition systems typically follow a structured sequence comprising four critical stages. Based on an extensive review of existing literature, these stages include data acquisition, data preprocessing, feature extraction, and gesture classification. Initially, hand gesture data are acquired through either sensor-based technologies or computer vision methodologies. Subsequently, data preprocessing is conducted to enhance image quality and standardize inputs, thereby facilitating more robust recognition performance. The next phase, feature extraction, involves the identification of salient characteristics such as hand shape, orientation, and motion trajectories. Finally, gesture classification is performed using machine learning or deep learning algorithms to accurately categorize the gestures.

Regarding data acquisition, two primary methodologies have been predominantly employed: sensor-based and vision-based approaches. Sensor-based acquisition relies on wearable gloves embedded with electromechanical sensors to capture detailed hand movements. While this method yields high precision, it is constrained by several factors, including the requirement for expensive, specialized hardware, user discomfort associated with wearing sensors, and limited scalability due to individual variations in hand morphology necessitating device recalibration. In contrast, vision-based acquisition utilizes camera systems to detect and interpret hand gestures without additional hardware, offering a more accessible and cost-effective solution. However, this approach is challenged by variations in hand appearance (e.g., differences in skin tone and hand size), susceptibility to background clutter and varying lighting conditions, and the complexity of accurately capturing dynamic gestures.

To enhance the reliability and accuracy of vision-based hand gesture recognition, multiple preprocessing and feature extraction techniques are incorporated. Gaussian blur filtering is applied to suppress image noise and enhance the clarity of hand gestures. Background normalization techniques are employed to maintain a uniform background environment, thereby mitigating the dependency on skin color-based segmentation. Additionally, hand landmark detection is performed using the MediaPipe Hand Tracking framework, enabling the extraction of 21 anatomically significant hand keypoints, which facilitates precise feature representation for downstream tasks.

For the classification of gestures, various machine learning and deep learning models have been investigated. Convolutional Neural Networks (CNNs) are frequently utilized for feature extraction tasks but are computationally intensive. Recurrent Neural Networks (RNNs) offer the advantage of modeling sequential data; however, they are susceptible to issues such as vanishing gradients, which impair long-term dependency learning. Long Short-Term Memory (LSTM) networks mitigate these limitations by effectively capturing temporal dependencies, making them particularly suitable for real-time applications such as Indian Sign Language (ISL) recognition. In the present study, an LSTM-based deep learning model was developed and implemented using TensorFlow and Keras. The model was trained on a custom ISL dataset and optimized to manage variations in lighting, hand orientations, and movement velocities.

Although substantial progress has been made in the domain of gesture recognition, existing systems are often constrained by high costs, hardware dependencies, and limited scalability. To address these challenges, the proposed system eliminates the need for specialized sensor hardware by utilizing a standard webcam for data acquisition.

Furthermore, the system enhances accessibility by providing an open-source, low-cost framework suitable for a broader range of users. By integrating advanced deep learning methodologies, specifically LSTM networks, with state-of-the-art computer vision techniques such as OpenCV and MediaPipe, the proposed approach achieves high accuracy and real-time gesture recognition performance, representing a significant advancement over conventional methodologies.

TECHNOLOGIES AND TOOLS USED

The development of a real-time Indian Sign Language (ISL) recognition system necessitates the integration of machine learning, deep learning, computer vision, and software development frameworks. The proposed implementation employs the following key components:

Programming Environment: Python was adopted as the primary language due to its extensive ecosystem supporting machine learning, image processing, and real-time application development.

Machine Learning and Deep Learning Frameworks: TensorFlow and Keras were utilized for constructing and training the LSTM-based deep learning model. Scikit-learn facilitated data preprocessing, feature extraction, and performance evaluation through metrics such as accuracy, precision, recall, and F1-score.

Computer Vision and Image Processing: OpenCV enabled webcam-based video capture, image preprocessing (Gaussian blur, thresholding, edge detection), and hand contour detection. MediaPipe enhanced recognition accuracy through real-time tracking of 21 hand landmarks and trajectory extraction.

Neural Network Architecture: An LSTM-based sequential deep learning model was implemented to capture temporal dependencies in gesture sequences, providing robust classification capabilities.

Dataset and Data Handling: The Kaggle ISL dataset (42,000 images across 36 classes) was employed for model training and validation. NumPy and Pandas managed data operations, while Matplotlib and Seaborn facilitated visualization of learning curves and dataset characteristics.

Development Environment: Visual Studio Code served as the primary IDE, supplemented by Jupyter Notebook for experimental analyses. Version control was maintained via Git and GitHub.

System Requirements: Minimum hardware specifications included an Intel Core i5 (10th Gen) CPU, NVIDIA GTX 980 GPU, 8GB RAM, and a webcam. Essential software dependencies included Python 3.x, TensorFlow, Keras, OpenCV, MediaPipe, NumPy, Pandas, Scikit-learn, Matplotlib, and Seaborn.

Deployment and Future Work: The system achieves real-time gesture recognition via webcam input. Future directions include cloud/edge deployment (AWS, Google Cloud), mobile platform integration, and NLP-based expansion to full-sentence translation capabilities.

MODEL ANALYSIS AND RESULT

The developed model for Indian Sign Language (ISL) gesture recognition was trained utilizing deep learning methodologies to achieve high accuracy in real-time gesture detection. A vision-based approach was adopted, employing a deep learning architecture specifically optimized for gesture classification tasks. The training dataset comprised a custom ISL dataset encompassing alphabets (A–Z), digits (0–9), and frequently used ISL gestures to ensure comprehensive system performance.

A. Model Training and Optimization

The model architecture was based on Long Short-Term Memory (LSTM) networks, selected for their superior ability to process sequential data such as temporal hand gesture movements. Supervised learning techniques were employed, leveraging labeled ISL datasets, while data augmentation strategies were incorporated to increase the diversity and robustness of training samples. Real-time hand tracking was facilitated through MediaPipe Hand Landmark Detection, enabling the extraction of 21 keypoints from the hand to serve as dynamic input features.

B. Model Performance Metrics

The performance of the trained model was systematically evaluated using standard classification metrics. Accuracy was measured to assess the overall correctness of predictions, while precision and recall were used to quantify the model's effectiveness in handling false positives and false negatives, respectively. The F1-score provided a balanced evaluation metric between precision and recall. The model achieved a training accuracy of 90.3% and a validation accuracy of 85.7%. In real-time testing scenarios, the gesture recognition accuracy ranged between 75–80%. The reduction in real-time performance was attributed primarily to environmental factors such as lighting variations, background noise, and inconsistencies in user hand positioning during live detection.

C. Transfer Learning for Model Optimization

To further enhance system performance and reduce computational complexity, transfer learning methodologies were employed. Transfer learning involves leveraging a model pre-trained on a related task, followed by fine-tuning it on the target ISL dataset. This approach enables faster convergence and improved classification accuracy. A pre-trained hand detection model was adapted for ISL-specific gesture classification through careful model re-training and optimization.

D. LSTM-Based Sequential Model Architecture

The final model architecture integrated an LSTM-based sequential learning structure. Input features were extracted from real-time video streams using OpenCV and MediaPipe frameworks. The model architecture comprised convolutional layers for initial feature extraction, LSTM layers for modeling temporal dependencies across sequential gestures, and fully connected dense layers for final multi-class classification. The model optimization utilized the Adam optimizer with adaptive learning rate tuning, and the categorical cross-entropy loss function was employed to handle multi-class classification tasks effectively.

E. Real-Time Testing and Observed Challenges

The system was deployed and tested in real-world environments to validate its performance under practical conditions. Several challenges were identified during real-time evaluation, including the impact of lighting variability on recognition accuracy, hand occlusions causing reduced classification confidence, and variations in user hand positioning leading to inconsistent detection outcomes. These challenges highlight the need for further model robustness enhancements and adaptive preprocessing techniques to improve real-time gesture recognition reliability.

6. Snapshots of system with brief detail of each

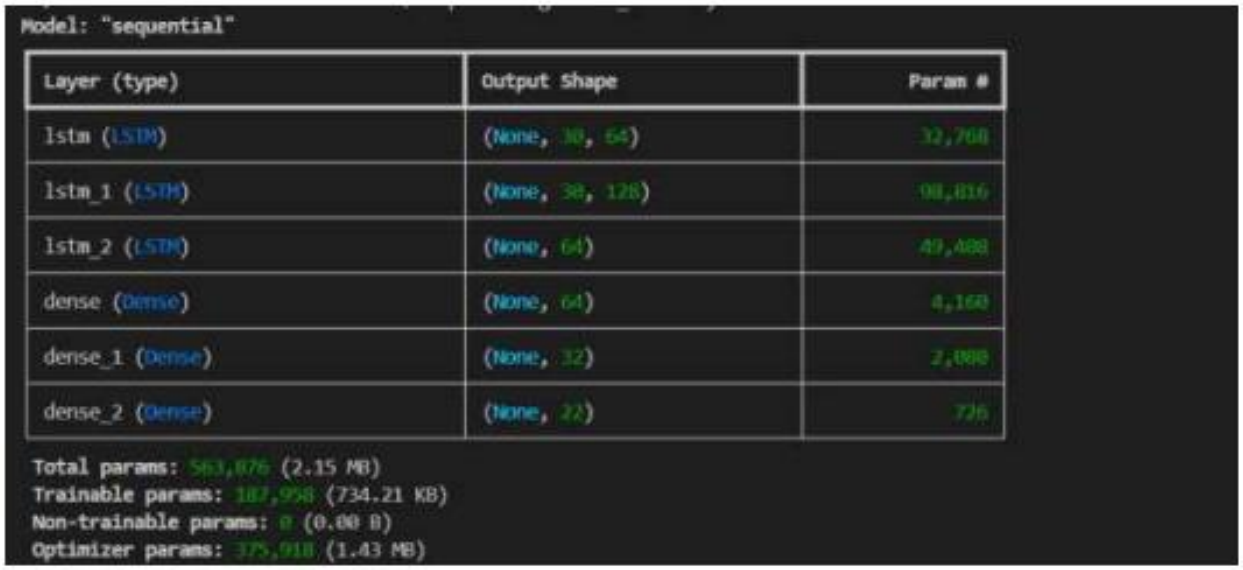

Figure V.1 Training of alphabets

Figure V.2 Detection of Alphabets

CONCLUSION

The integration of machine learning and computer vision techniques into sign language recognition systems presents significant potential for enhancing accessibility and facilitating communication for individuals with hearing and speech impairments. By leveraging advanced deep learning architectures, the proposed system offers an efficient, real-time solution for static hand gesture interpretation, thereby contributing to more inclusive communication frameworks.

The developed system successfully demonstrates accurate recognition of static hand gestures; however, it currently encounters limitations in dynamic gesture recognition and continuous speech sequence interpretation. Performance fluctuations were observed due to variations in hand positioning, illumination conditions, and ambient background noise. Future work will focus on expanding the dataset, optimizing real-time hand tracking algorithms, and incorporating dynamic gesture recognition capabilities. Moreover, the integration of Natural Language Processing (NLP) techniques is planned to enable full-sentence recognition and translation, thereby enhancing the system's contextual understanding and conversational fluency.

This research proposes a cost-effective, vision-based alternative to traditional sensor-dependent systems, thereby eliminating the reliance on expensive hardware such as sensor gloves. Continued advancements in machine learning and artificial intelligence are anticipated to further refine the system's accuracy, robustness, and usability, paving the way for broader deployment in real-world environments and contributing to a more inclusive society for the deaf and speech-impaired community.

REFERENCES

[1]. Nature. (2023). *Sign language recognition using deep learning*. [online] Available at: https://www.nature.com/articles/s41598-023-43852-x [Accessed 26 Mar. 2025].
[2]. Arikeri, P. (n.d.). *Indian Sign Language (ISL) Dataset*. [online] Kaggle. Available at: https://www.kaggle.com/datasets/prathumarikeri/indian-sign-language-isl/code [Accessed 26 Mar. 2025].
[3]. 20it105. (n.d.). *Sign Language Recognition Using Python*. [online] Medium. Available at: https://medium.com/@20it105/sign-language-recognition-using-python-74ef7ea43181#:~:text=The%20Sign%20Language%20Recognition%20project,image%20processing%2C%20and%20machine%20learning [Accessed 26 Mar. 2025].
[4]. Data Flair. (n.d.). *Sign Language Recognition Using Python and OpenCV*. [online] Available at: https://data-flair.training/blogs/sign-language-recognition-python-ml-opencv/ [Accessed 26 Mar. 2025].
[5]. GeeksforGeeks. (n.d.). *Sign Language Recognition System Using TensorFlow in Python*. [online] Available at: https://www.geeksforgeeks.org/sign-language-recognition-system-using-tensorflow-in-python/ [Accessed 26 Mar. 2025].
[6]. Jayar, E. (n.d.). *American Sign Language Recognition and Processing*. [online] University of Toronto. Available at: https://www.eecg.utoronto.ca/~jayar/mie324/asl.pdf [Accessed 26 Mar. 2025].