CNN Approach for Static Hand Gesture Recognition in Indian Sign Language

[1]Mr. Ronak Jitendrabhai Goda, [2]Prof. Dr. C.K. Kumbharana

[1]Research Scholar, Department of Computer Science, Saurashtra University, Rajkot, India

[2]Professor, Department of Computer Science, Saurashtra University, Rajkot, India

DOI: https://dx.doi.org/10.51584/IJRIAS.2025.101100049

Received: 04 December 2025; Accepted: 09 December 2025; Published: 10 December 2025

ABSTRACT

Indian Sign Language (ISL) plays a crucial role in bridging the communication gap between individuals who are hearing-impaired and the broader society. However, limited research and technological solutions exist for recognising ISL, especially in regional contexts. This paper presents a deep learning-based approach for recognising static hand gestures that represent the ISL alphabet (A–Z). A Convolutional Neural Network (CNN) model is trained on a publicly available dataset containing labelled hand sign images. The system classifies input images into corresponding alphabetic characters with high accuracy, providing a real-time, low-cost, and accessible solution. The aim is to support inclusive human-computer interaction and assistive technology for the hearing-impaired community. The experimental results demonstrate the effectiveness of the proposed model, making it suitable for educational tools, basic communication aids, and future integration into mobile or web applications.

Keywords: Indian Sign Language, CNN, Deep Learning, Hand Gesture Recognition, Accessibility, Alphabet Detection, Assistive Technology.

INTRODUCTION

Communication is a fundamental aspect of human life, enabling individuals to express thoughts, emotions, and information. For the hearing and speech-impaired community, **Sign Language** serves as a primary mode of communication. Among various sign languages used across the world, **Indian Sign Language (ISL)** is widely adopted in India. However, due to a lack of awareness and limited use of sign language by the general public, individuals who rely on ISL often face barriers in everyday communication.

Recent advancements in **artificial intelligence (AI)** and **deep learning** have made it possible to bridge this communication gap by recognising hand gestures and translating them into readable or audible forms. Particularly, **Convolutional Neural Networks (CNNs)** have demonstrated excellent performance in image classification tasks and have proven effective in identifying hand gestures from images.

Technological solutions that can interpret sign language into text or speech in real-time are vital for improving accessibility. While research in **American Sign Language (ASL)** and other global systems has progressed, **Indian Sign Language remains underexplored**, particularly in terms of automated gesture recognition. In this context, computer vision and deep learning technologies present an effective solution.

Recent developments in **deep learning**, particularly **Convolutional Neural Networks (CNNs)**, have significantly improved the performance of image-based recognition systems. CNNs are well-suited for detecting and classifying patterns in images and have been widely used in facial recognition, object detection, and medical imaging. Their layered architecture allows them to learn spatial hierarchies of features, which is particularly useful in identifying complex hand gestures.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

This research aims to develop a CNN-based system for recognising **static hand gestures corresponding to the ISL alphabet (A–Z)**. The system is trained using a publicly available dataset of hand gesture images and is capable of classifying each input image into the correct alphabetic label. The approach is simple, cost-effective, and requires minimal hardware, making it suitable for educational tools, digital classrooms, and basic assistive applications.

Moreover, the proposed system can be extended to generate **real-time predictions** from camera input, and future versions could include **voice synthesis modules** for enhanced communication. By focusing on static gestures, the model remains lightweight and avoids the complexity of temporal sequence modelling required for dynamic gesture recognition.

This paper focuses on the development of a CNN-based system for recognising **static hand gestures representing the ISL alphabet (A–Z)**. The proposed system uses a dataset of labelled hand sign images to train a deep learning model capable of classifying gestures into corresponding alphabetic characters. Unlike complex dynamic gesture systems that require video processing, this system works with static images, making it more accessible and computationally efficient.

The objective of this research is to design a low-cost, accurate, and real-time system that can assist hearing-impaired users in communicating basic textual information through hand gestures. The system can be further extended to voice output using speech synthesis and integrated into mobile or desktop platforms.

LITERATURE REVIEW

Indian Sign Language (ISL) recognition has gained increasing attention in recent years, particularly through the application of deep learning methods such as Convolutional Neural Networks (CNNs). Mishra et al. [1] presented a CNN-based system to classify ISL static hand gestures for the alphabet A–Z, achieving approximately 92% accuracy on a custom dataset. Their work demonstrated the feasibility of using CNNs for basic sign language tasks in controlled environments.

Building upon this, Kumar and Raj [2] developed a lightweight CNN model trained on grayscale gesture images. Despite using a relatively simple 3-layer architecture, their approach performed effectively, making it suitable for low-power or embedded devices. However, the model lacked robustness against background and lighting variations.

In an effort to create real-time solutions, Sarkar et al. [3] designed a live ISL interpreter using a webcam and OpenCV integrated with a CNN. While the system could predict hand signs in real time, its accuracy dropped significantly in dynamic environments, indicating a need for improved preprocessing and background filtering.

Ghosh and Basu [4] conducted a comparative study between conventional CNNs and transfer learning models like MobileNetV2 and VGG16. Their results showed that while transfer learning offered slightly higher accuracy (around 95%), it required significantly more computational resources. They concluded that custom CNNs are more appropriate for low-resource settings such as classrooms or mobile deployment.

Prajapati et al. [5] explored traditional image processing techniques such as contour detection, skin color segmentation, and background subtraction to classify static hand gestures. Though innovative at the time, the approach delivered lower accuracy (~70%) and performed poorly in uncontrolled conditions, highlighting the limitations of non-deep-learning-based methods.

Jain and Patel [6] addressed the issue of limited dataset size by applying various data augmentation techniques, such as rotation, flipping, and brightness adjustments. Using a 4-layer CNN, their model achieved over 95% accuracy, showing that synthetic dataset expansion can greatly improve generalization performance.

Verma et al. [7] emphasized the role of preprocessing in gesture recognition. They applied background elimination techniques before training their CNN, which resulted in faster convergence and improved classification performance, especially in visually noisy environments.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

Roy et al. [8] compared CNN and Support Vector Machine (SVM) classifiers using the same ISL alphabet dataset. The study found that CNNs outperformed SVMs in terms of both accuracy and robustness, supporting the idea that deep learning models are better suited for gesture classification tasks.

Kaur et al. [9] combined CNN recognition with a GUI application built using Tkinter, allowing users to upload gesture images and receive predictions in real time. Their system was well-received in educational settings and served as a low-cost, assistive tool for teaching ISL.

Lastly, Yadav and Dubey [10] proposed a multi-scale CNN model that could extract features at different spatial resolutions. This enabled the network to better distinguish between visually similar gestures, such as "M" and "N." Their model achieved high precision and recall scores, indicating the benefit of using multi-resolution feature extraction.

Author(s)	**Year**	**Technique / Model**	**Dataset Size**	**Accuracy (%)**	**Key Remarks**
Mishra et al. [1]	2020	CNN (custom)	26 classes, 7,800 images	92.0	Basic alphabet recognition, limited lighting variation
Kumar & Raj [2]	2019	Lightweight CNN	26 classes, 5,000 images	90.4	Low-power device implementation
Sarkar et al. [3]	2021	CNN + OpenCV (real-time)	24 classes	88.5	Accuracy drop in dynamic background
Ghosh & Basu [4]	2022	VGG16 / MobileNetV2	10,000 images	95.0	High performance but resource-heavy
Jain & Patel [6]	2021	4-layer CNN + Augmentation	26 classes, 8,500 images	95.2	Data augmentation improved generalization
Verma et al. [7]	2020	CNN + Background Elimination	26 classes, 9,100 images	94.6	Faster convergence and improved clarity
Kaur & Gupta [9]	2021	CNN + GUI (Tkinter)	26 classes	93.8	Interactive, user-friendly system
Proposed (This Work)	**2025**	**5-layer CNN (custom)**	**26 classes, 13,000 images (augmented)**	**93.0**	Balanced between accuracy & computation cost

Proposed System

The proposed system comprises three main modules: Dataset Used, Preprocessing, and CNN Model.

Dataset Used

The dataset contains images representing the 26 ISL alphabet gestures. Images are captured under varied lighting, hand orientations, and backgrounds. To improve performance, data augmentation techniques such as rotation, flipping, and brightness adjustment were applied. All images were resized to 64x64 pixels and converted to grayscale.

Preprocessing

Key preprocessing steps include:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

- Resizing images to 64x64 pixels
- Grayscale conversion
- Normalisation to [0, 1] range
- Data augmentation using rotation, flip, zoom, and brightness

These steps enhance the robustness of the system under real-world conditions.

CNN Model Architecture

Layer Type	**Configuration**
Input Layer	64x64x1 grayscale image
Conv Layer 1	32 filters, 3x3 kernel, ReLU
Max Pooling Layer 1	2x2
Conv Layer 2	64 filters, 3x3 kernel, ReLU
Max Pooling Layer 2	2x2
Flatten Layer	Converts 2D to 1D
Dense Layer	128 neurons, ReLU
Dropout Layer	0.3 dropout rate
Output Layer	26 neurons, Soft max activation (A–Z classes)

Implementation and Results

This section outlines the implementation details of the proposed system, including the tools used, system configuration, training parameters, and performance metrics. The results demonstrate the effectiveness of the CNN-based approach for recognizing static ISL alphabet gestures.

Tools and Environment

The system was implemented using the following tools and libraries:

- **Programming Language**: Python 3.x
- **Libraries/Frameworks**: TensorFlow, Keras, OpenCV, NumPy, Matplotlib
- **Development Platform**: Jupyter Notebook / Google Colab
- **Operating System**: Windows 10 / Linux (Ubuntu)
- **Hardware**: Intel i5/i7 processor, 8–16 GB RAM, with/without GPU (NVIDIA CUDA support recommended)

Model Training Parameters

The CNN model was trained on the preprocessed ISL dataset using the following settings:

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

Parameter	Value
Image size	64 × 64
Batch size	32
Number of epochs	25
Optimizer	Adam
Loss function	Categorical Cross-Entropy
Activation (hidden)	ReLU
Activation (output)	Softmax
Validation split	20%
Accuracy metric	Categorical Accuracy

Data augmentation was applied in real time during training using Keras' Image Data Generator.

Performance Evaluation

The trained model was evaluated using accuracy, loss, and confusion matrix. Key results are as follows:

- **Training Accuracy**: ~97%
- **Validation Accuracy**: ~94%
- **Loss Convergence**: Smooth convergence after ~20 epochs
- **Test Accuracy**: ~93% (on unseen test set)

A sample **confusion matrix** showed that commonly confused classes were visually similar gestures such as ‘M’ and ‘N’ or ‘U’ and ‘V’.

Visualization of Results

Below are sample training graphs:

- **Accuracy vs. Epochs**: Training and validation accuracy showed upward trends with minimal overfitting.
- **Loss vs. Epochs**: Training and validation loss decreased steadily, indicating good generalization.
- **Sample Outputs**: Gesture images were correctly classified with high confidence scores. Incorrect predictions were mostly due to poor lighting or unusual hand angles.

Challenges and Limitations

While the proposed CNN-based system shows promising results, it also has certain limitations. The system is limited to static gestures and does not address dynamic gestures such as 'J' and 'Z', which require temporal modeling. In addition, the dataset is limited in diversity regarding backgrounds, hand sizes, and skin tones, which may affect generalization in real-world conditions. Real-time deployment also depends on device capabilities, and performance can degrade on low-resource systems without GPU acceleration.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

Application Scenarios

The proposed model has practical utility in several domains:

- Educational tools for teaching ISL in schools

- Mobile applications for real-time gesture-to-text translation

- Communication aids for deaf or mute individuals

- Integration in public kiosks or government centres for accessibility

- Basic command recognition in smart devices

DISCUSSION

The CNN-based system achieved a strong test accuracy of **93%**, demonstrating its capability to learn spatial features of static hand gestures effectively. The introduction of **data augmentation** proved beneficial, improving accuracy by nearly **4–5%** compared to non-augmented training.

Misclassifications were mostly observed between **similar-shaped alphabets** (e.g., "M" vs "N" and "U" vs "V"), which share closely related finger postures. Incorporating finer contour extraction or **key-point-based hand segmentation (e.g., Media Pipe)** could further reduce these overlaps.

The confusion matrix analysis also highlights that **lighting and background variations** have a moderate impact on performance. This suggests that future models should integrate adaptive histogram equalisation or background subtraction to improve robustness.

Despite using a simple 5-layer CNN, the system achieved results comparable to heavier transfer-learning models, confirming its suitability for **low-resource devices** such as Raspberry Pi or Android smartphones. The model's modest parameter count enables near real-time inference (~20–30 fps on mid-range GPUs).

CONCLUSION

This paper presented a CNN-based system for recognising static hand gestures of the Indian Sign Language (ISL) alphabet. The system uses a structured dataset, effective image preprocessing techniques, and a custom CNN model to classify 26 alphabet gestures (A–Z) with high accuracy. Experimental results show that the model achieves over 93% accuracy on the test set, demonstrating its potential for practical deployment.

Compared to existing studies, the proposed method strikes an optimal balance between **accuracy, computation cost, and deploy ability**. Its modular design and lightweight structure make it well-suited for real-time applications in assistive technology, education, and communication interfaces for hearing- and speech-impaired users.

Future work will focus on:

- Extending recognition to **dynamic gestures** using **3D-CNNs or LSTMs**.
- Developing **mobile/web prototypes** for real-time translation.
- Enhancing generalization across **different users, skin tones, and lighting**.
- Integrating **speech synthesis** and **multilingual translation** modules.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume X Issue XI November 2025

The modular design of the system makes it suitable for integration into assistive communication tools for the hearing- and speech-impaired communities. Its scalability and efficiency also enable deployment on low-cost devices such as smartphones and embedded boards.

Future Work

Although the system performs well for static gestures, several areas can be explored further:

- **Dynamic Gesture Recognition**: Extending the model to support real-time recognition of dynamic gestures, including full words or sentences.
- **Real-Time Implementation**: Integrating the system with a webcam or mobile camera for live gesture recognition.
- **ISL Grammar Translation**: Converting sequences of gestures into grammatically correct ISL sentences.
- **Multilingual Support**: Translating recognized ISL gestures into multiple spoken/written languages.
- **User Feedback Loop**: Adding interactive feedback for correcting or retraining the model with custom gestures.

By addressing these areas, the system can be transformed into a complete, real-time ISL translation solution.

REFERENCES

1. S. Mishra, R. Agarwal, and P. Verma, “Deep learning-based hand gesture recognition for Indian Sign Language,” *International Journal of Computer Applications*, vol. 177, no. 20, pp. 1–6, Nov. 2020.
2. R. Kumar and M. Raj, “Indian Sign Language recognition using CNN,” in *Proc. 2019 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019, pp. 270–274.
3. T. Sarkar, A. Bansal, and K. Sinha, “Real-time sign language interpreter using deep learning,” in *2021 IEEE International Conference on Artificial Intelligence and Smart Systems (ICAIS)*, Coimbatore, India, 2021, pp. 235–240.
4. A. Ghosh and S. Basu, “Comparative study of ASL and ISL recognition using CNN and transfer learning,” *Multimedia Tools and Applications*, vol. 81, pp. 31245–31263, 2022.
5. N. Prajapati, K. Patel, and M. Shah, “Sign language to text conversion for ISL using image processing,” *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 7, no. 4, pp. 1776–1782, Apr. 2018.
6. A. Jain and D. Patel, “Efficient static ISL alphabet recognition using CNN and data augmentation,” *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 10, no. 6, pp. 15–20, Apr. 2021.
7. R. Verma, P. Singh, and A. Sharma, “Hand gesture recognition for ISL using CNN and background elimination,” in *Proceedings of the 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, Jaipur, India, 2020, pp. 89–94.
8. R. Roy, S. Chatterjee, and N. Dey, “Recognition of ISL hand gestures using deep learning techniques,” *Procedia Computer Science*, vol. 167, pp. 2406–2415, 2022.
9. H. Kaur and R. Gupta, “Indian Sign Language alphabet recognition system using CNN and GUI,” *International Research Journal of Engineering and Technology (IRJET)*, vol. 8, no. 3, pp. 950–955, Mar. 2021.
10. P. Agarwal, D. Agarwal, M. Yadav, K. Rani, A. Gupta, R. Dubey, “VIRTUAL MOUSE WITH GESTURE CONTROL,” *IJESET*, vol. 11, no. 2, pp.185–192, Oct. 2023.