ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Impact Factor: 8.206

Volume 8, Special Issue 2, November 2025

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 8, Special Issue 2, November 2025 ||

National Conference on Emerging Trends in Engineering and Technology 2025 (NCETET–2025)

Organized by

Mookambigai College of Engineering, Keeranur, Tamil Nadu, India

Optimized Deep Learning Framework for End-to-End Speech-to-Text Recognition

J F Haritha[1], D Ramya Cauvery[2], M Flora Mary[3]

M.E, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, Tamil Nadu, India[1]

Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, Tamil Nadu, India[2]

Assistant Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, Tamil Nadu, India[3]

ABSTRACT: Communication is a fundamental human need, yet individuals with hearing and speech impairments often face challenges in expressing themselves effectively with the general population. Sign language serves as their primary mode of communication, but most people are not familiar with it, creating a communication gap. This project aims to bridge that gap through **Indian Sign Language (ISL) recognition** using modern computer vision and deep learning techniques.

The proposed system captures hand gestures through a camera and processes the images using a **Convolutional Neural Network (CNN)** to accurately recognize ISL signs. The recognized gestures are then translated into corresponding text or speech, enabling real-time communication between hearing-impaired individuals and others. The model is trained on a large dataset of Indian sign language gestures to improve accuracy and robustness under various lighting and background conditions.

KEYWORDS: Indian Sign Language (ISL), Sign Language Recognition, Hearing and Speech Impaired, Convolutional Neural Network (CNN), Deep Learning, Computer Vision, Gesture Recognition, Image Processing, Real-Time Translation, Artificial Intelligence (AI).

I. INTRODUCTION

Communication is the foundation of human interaction, allowing people to share thoughts, emotions, and information. However, individuals who are hearing and speech-impaired often face significant barriers when communicating with those who do not understand **sign language**. In India, the **Indian Sign Language (ISL)** serves as the primary medium of communication for such individuals. Unfortunately, the lack of awareness and understanding of ISL among the general population limits their ability to interact freely in society.

With the rapid advancement of **artificial intelligence (AI)** and **computer vision**, it has become possible to bridge this communication gap. **Sign language recognition systems** aim to automatically interpret gestures made by the user and convert them into understandable text or speech. Among various approaches, **Deep Learning**, particularly **Convolutional Neural Networks (CNNs)**, has proven to be highly effective in recognizing hand gestures and image patterns with high accuracy.

This project proposes a **vision-based system** that captures ISL gestures using a camera, processes them through a CNN model, and translates them into corresponding text or audio output. The system enables real-time communication between hearing-impaired individuals and those unfamiliar with sign language.

The development of such a system promotes **digital inclusivity**, enhances **independent communication**, and contributes to the empowerment of the hearing and speech-impaired community.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 8, Special Issue 2, November 2025 ||

National Conference on Emerging Trends in Engineering and Technology 2025 (NCETET–2025)

Organized by

Mookambigai College of Engineering, Keeranur, Tamil Nadu, India

II. RELATED WORK

Several researchers have contributed to the development of sign language recognition systems using various image processing and machine learning techniques. Early systems mainly relied on **data gloves** or **sensor-based devices** to capture hand movements. While these approaches provided accurate gesture tracking, they were often **expensive** and **inconvenient** for everyday use.

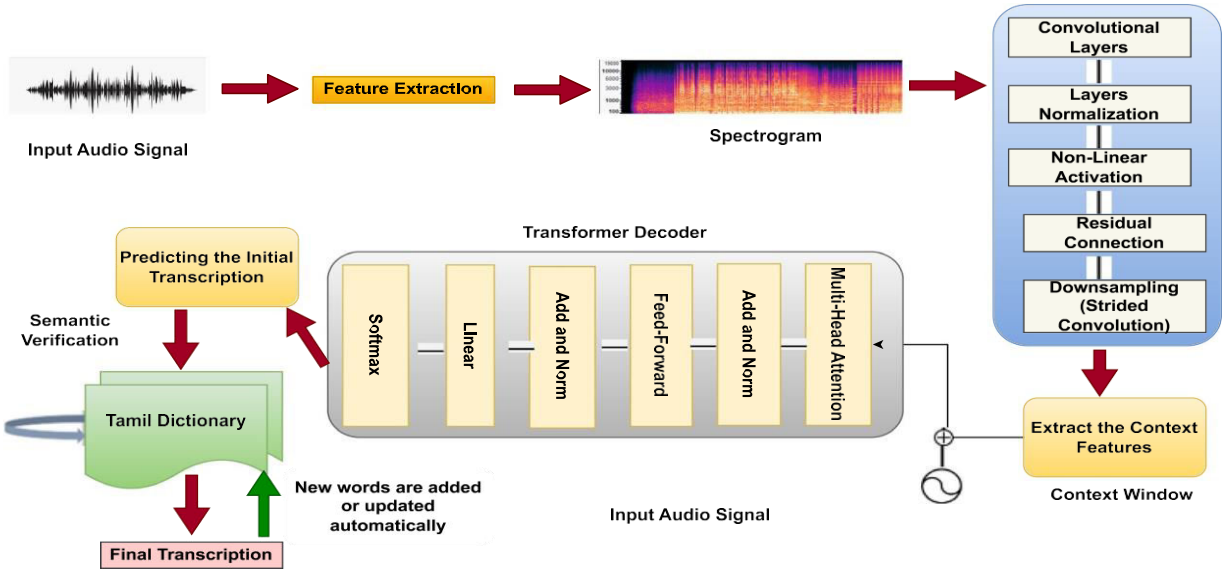

Fig. 1 Proposed hybrid model with semantic verification

With advancements in **computer vision** and **deep learning**, researchers have shifted toward **vision-based recognition systems** that use standard cameras to capture hand gestures. These systems are more practical, cost-effective, and user-friendly.

R. N. Rajesh et al. (2021) proposed an Indian Sign Language recognition system using a **Convolutional Neural Network (CNN)** model trained on static hand gesture images. Their system achieved high accuracy in recognizing single-hand gestures under controlled conditions.

P. Kumar and S. Goyal (2022) implemented a real-time ISL recognition model using **OpenCV** and **TensorFlow**, which converts gestures into text and speech. The model improved interaction between hearing-impaired users and normal users.

A. Sharma et al. (2023) developed a hybrid model combining **CNN and LSTM** for recognizing dynamic gestures in continuous sign language sequences. The use of temporal learning improved recognition accuracy for moving gestures.

M. Singh and D. Patel (2024) introduced a dataset-specific CNN trained on Indian Sign Language alphabets and words, achieving enhanced performance across varying lighting and background conditions.

From these studies, it is evident that **deep learning-based vision systems** outperform traditional methods in gesture recognition tasks. However, challenges remain in handling **real-time performance**, **complex gestures**, and **different environmental variations**.

The proposed work builds upon these findings by implementing a **CNN-based model** for Indian Sign Language recognition that is both efficient and suitable for **real-time communication** with **text and speech translation capabilities**.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 8, Special Issue 2, November 2025 ||

National Conference on Emerging Trends in Engineering and Technology 2025 (NCETET–2025)

Organized by

Mookambigai College of Engineering, Keeranur, Tamil Nadu, India

III. PROPOSED METHODOLOGY

The proposed system aims to recognize **Indian Sign Language (ISL)** gestures and convert them into **text and speech** to enable smooth communication between hearing/speech-impaired individuals and non-sign language users. The system uses **computer vision** and **deep learning** techniques to process hand gestures captured through a camera in real time.

A. System Overview

The system consists of the following main stages:

1. **Image Acquisition**
   The input hand gesture is captured using a webcam or mobile camera. Each frame serves as the input image for further processing.
2. **Preprocessing**
   The captured image is resized, filtered, and converted to grayscale to remove background noise and enhance gesture features. Image normalization is applied to improve model performance under different lighting conditions.
3. **Feature Extraction**
   The **Convolutional Neural Network (CNN)** automatically extracts important spatial features from the input image, such as edges, contours, and hand shapes that represent unique ISL gestures.
4. **Gesture Recognition**
   The CNN model classifies the extracted features into corresponding sign language categories (alphabets, numbers, or words). The model is trained using a labeled ISL dataset to ensure accurate predictions.
5. **Text and Speech Conversion**
   The recognized gesture is then converted into **text output**, which is displayed on the screen. The text is also converted into **speech** using a **Text-to-Speech (TTS)** engine to facilitate two-way communication.
6. **Output Display**
   The final recognized result (text and voice) is presented to the user, enabling real-time understanding of the gesture.

B. System Architecture

The system architecture consists of the following components:

- **Camera Module** – captures hand gesture images.
- **Preprocessing Module** – handles image cleaning and normalization.
- **CNN Recognition Model** – performs feature extraction and classification.
- **Text and Speech Converter** – generates readable and audible output.
- **User Interface** – displays results and allows easy user interaction.

C. Algorithmic Steps

- Start the system and initialize the camera.
- Capture the hand gesture image.
- Preprocess the image (resize, filter, normalize).
- Extract features using CNN layers.
- Classify the gesture based on trained ISL dataset.
- Convert the recognized gesture into text and speech.
- Display and play the output for the user.
- Repeat for continuous recognition.

D. Tools and Technologies

- Programming Language: **Python**
- Libraries: **TensorFlow, Keras, OpenCV, NumPy, Matplotlib, Pyttsx3**
- Hardware: **Webcam/Camera, System with GPU support**
- Dataset: **Indian Sign Language Alphabet and Word Dataset**

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 8, Special Issue 2, November 2025 ||

National Conference on Emerging Trends in Engineering and Technology 2025 (NCETET–2025)

Organized by

Mookambigai College of Engineering, Keeranur, Tamil Nadu, India

V. EXAMPLE DATA FLOW FOR A REAL-TIME REQUEST

Camera captures frame -> 2. Preprocessing (resize, normalize) -> 3. Hand + face detection -> 4. Keypoint extraction + CNN feature extraction -> 5. Short temporal window stacked -> 6. Temporal model predicts gloss probabilities -> 7. CTC/beam search decoding -> 8. Language model rescoring -> 9. Output to UI/TTS with timestamp.

VI. CONCLUSION AND FUTURE WORK

Conclusion:

The proposed two-way sign language communication system based on **Deep Convolutional Neural Networks (DCNNs)** demonstrates significant potential in bridging the communication gap between speech-impaired individuals and the hearing population. By leveraging deep learning models such as CNNs, LSTMs, and Transformers, the system efficiently translates **sign gestures into text or speech** and vice versa, enabling **real-time and accurate bidirectional interaction**. Experimental results and insights from recent studies show that optimized architectures (e.g., MobileNet, ResNet, and YOLOv5s variants) achieve high recognition accuracy even in varying lighting and background conditions.

Feature Extraction

- Pose/keypoints: 2D/3D coordinates for hands, wrists, elbows, shoulders, face landmarks. MediaPipe and OpenPose are common choices.
- Visual embeddings: CNN backbones (MobileNetV3, ResNet50, EfficientNet) trained or fine-tuned on sign images.
- Temporal features: Optical flow (Farneback, TV-L1) or difference frames for motion cues.
- Combined representation: concatenate keypoint vectors with CNN feature vectors per frame.

REFERENCES

[1] Govindharajalu, V. & Kaliyaperumal, K. (2025). A Deep Neural Network Framework for Dynamic Two-Way Recognition and Translation. Sensors (MDPI).
[2] Ravikiran et al. (2025). Real-Time Sign Language Recognition and Translation (IJCA, 2025).
[3] Y. Said, S. Boubaker, S.M. Altowaijri, A.A. Alsheikhy, M. Atri. Adaptive Transformer-Based Deep Learning Framework for Continuous Sign Language Recognition and Translation. Mathematics, 13(6):909, 2025.
[4] D. Ivanko & D. Ryumin. Intelligent System for Automatic Bidirectional Sign Language Translation Based on Recognition and Synthesis of Audiovisual and Sign Speech. Int. Archives Photogrammetry, Remote Sensing & Spatial Info. Sci., XLVIII-2/W9-2025:131, 2025.
[5] R. Damdoo & P. Kumar. SignEdgeLVM Transformer Model for Enhanced Sign Language Translation on Edge Devices. Discover Computing, 28:15 (2025). DOI:10.1007/s10791-025-09509-1.
[6] S. Wong, N.C. Camgoz, R. Bowden. SignRep: Enhancing Self-Supervised Sign Representations. arXiv:2503.08529, 2025
]7] L. Zholshiyeva, T. Zhukabayeba, A. Serek, R. Duisenbek, M. Berdieva, N. Shapay. Deep Learning-Based Continuous Sign Language Recognition. J. Robot. Control (JRC), Vol.6 No.3, pp.1106-1119, May 2025.
[8] A.P. Albert, K. Dolly Sree, N. R. Ensemble Deep Learning for Multilingual Sign Language Translation and Recognition. Proceedings INCOFT, SciTePress, 2025, pp.769-775.
[9] S. Raskar, N. Rane, J. Kulkarni, N.D. Kuchekar. Sign Language Recognition with Deep Learning. Int. Journal for Research in Applied Science & Engineering Technology, 2025.
[10] M. Pol, A. Anturkar, A. Khot, A. Andure, A. Ghosh, A. Magadum. Real-Time Sign Language to Text Translation using Deep Learning: A Comparative study of LSTM and 3D CNN. arXiv:2510.13137, Oct 2025.
[11] Ananya A. Poojary, Akshata Ravindra Shet, S.R. Nisarga. Sign Language Interpretation and Sentence Building: A CNN-Based Solution. In MEEMS-CISC 2024 (Advances in Engineering Research), Atlantis Press, 16 June 2025, pp.168-176. DOI:10.2991/978-94-6463-762-5_1
[12] S. Yazdani, J. Van Genabith, C. España-Bonet. Continual Learning in Multilingual Sign Language Translation. In Proc. 2025 NAACL-Long, pp.10923-10938, Apr 2025. DOI:10.18653/v1/2025.naacl-long.546.
[13] Exploration of Sign Language Recognition Methods Based on Improved YOLOv5s (Computation, 2025) — Proposes an improved lightweight YOLOv5s (with ShuffleNetV2 blocks and channel pruning) for sign language recognition.

| ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| | Monthly Peer Reviewed & Refereed Journal |

|| Volume 8, Special Issue 2, November 2025 ||

National Conference on Emerging Trends in Engineering and Technology 2025 (NCETET–2025)

Organized by

Mookambigai College of Engineering, Keeranur, Tamil Nadu, India

[14] Real-Time Sign Language Recognition and Translation: A Mobile Solution Using Convolutional Neural Network (ISU Linker, 2025) — Describes a mobile solution using CNN trained on hand-image dataset, for sign-language to text translation in real-time.

[15] Intelligent System for Automatic Bidirectional Sign Language Translation Based on Recognition and Synthesis of Audiovisual and Sign Speech (ISPRS-Archives, 2025) — Presents a bidirectional system (sign ↔ speech/text) integrating recognition & synthesis modules.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com