# A UNIFIED DEEP LEARNING FRAMEWORK FOR INDIAN SIGN LANGUAGE INTERPRETATION AND SPEECH RECOGNITION

[1] Mr. N NARESH REDDY, [2] KANURI SOWMYA SREE, [3] THOTA DEEPAK SHETTY, [4] K. SRI HARI KRISHNA, [5] BALASANI VARSHITH

[1] *Assistant Professor, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering , Hyderabad, India.*

[2,3,4,5] *Students, Department of Computer Science & Engineering (Artificial Intelligence & Machine Learning), Malla Reddy College of Engineering, Hyderabad, India.*

***To Cite this Article***

*Mr. N Naresh Reddy, Kanuri Sowmya Sree, Thota Deepak Shetty, K. Sri Hari Krishna, Balasani Varshith, "A Unified Deep Learning Framework For Indian Sign Language Interpretation And Speech Recognition", Journal of Science Engineering Technology and Management Science, Vol. 02, Issue 11, November 2025,pp: 79-82, DOI:* http://doi.org/10.64771/jsetms.2025.v02.i11.pp77-82

*Submitted: 10-10-2025* *Accepted: 22-11-2025* *Published: 29-11-2025*

**ABSTRACT**

This work presents a unified deep learning–based Indian Sign Language (ISL) and speech recognition system designed to bridge communication gaps between hearing-impaired individuals and the general population [1], [7]**.** The proposed framework integrates image-based gesture recognition with natural speech-to-text conversion to enable seamless two-way interaction, following earlier efforts in sign-to-speech and gesture-to-text translation systems [1], [2], [4]**.** Using convolutional neural networks (CNN), feature extraction from hand gestures is performed, while recurrent and transformer-based models handle speech recognition tasks [11], [12]**.** The system aims to deliver high accuracy, robustness to noise, and real-time performance, improving upon traditional ISL recognition and glove-based solutions [2], [6], [9] and making it suitable for assistive communication applications [3], [5], [8]**.**

**Keywords:** Indian Sign Language, Speech Recognition, Deep Learning, CNN, Gesture Recognition, Assistive Technology



## I. INTRODUCTION

Indian Sign Language (ISL) is one of the primary modes of communication for the deaf and hard-of-hearing community [7], [12]. However, most individuals outside this community are not familiar with ISL, creating communication barriers in educational institutions, workplaces, public spaces, and healthcare facilities. Prior studies emphasize that automated sign language translation systems can enhance accessibility and inclusivity [1], [3], [4]. With recent advancements in computer vision and speech processing, deep learning has emerged as a powerful tool for interpreting gestures and spoken language [5], [9], [11]. Traditional ISL recognition systems relied heavily on handcrafted features, specialized gloves, or controlled environmental conditions, which limited real-world deployment [2], [6], [9]. Likewise, existing speech recognition frameworks often struggled with Indian accents, multilingual variations, and noisy environments, motivating the adoption of modern deep learning architectures [8],

[10], [12]. Integrating ISL recognition and speech-to-text translation into a unified model provides a two-way communication interface that is intuitive, efficient, and scalable [1], [4], [7].

The goal of the present work is to design a real-time ISL and speech recognition framework capable of high accuracy and robust performance. By leveraging CNNs for gesture feature extraction and transformer-based architectures for acoustic modeling, the system enhances recognition accuracy while maintaining low computational cost [5], [11], [12]. Such a system holds promise for deployment across educational, healthcare, and public-service domains, contributing to greater inclusivity and accessibility for the deaf community [3], [7], [8].

## II. LITERATURE SURVEY

**Author 1: Patel & Sharma (Deep Learning-Based ISL Gesture Classification)**

Patel and Sharma explored the use of convolutional neural networks for classifying static ISL gestures. Their work demonstrated that CNNs outperform traditional feature-engineering approaches by learning hierarchical representations directly from images. This reduced manual effort and improved classification accuracy across diverse gesture patterns. Their research focused on building large annotated datasets with variations in lighting, angles, and hand shapes. They found that data augmentation greatly improved generalization and robustness. Additionally, they showed that deeper CNN architectures like VGG and ResNet delivered higher accuracy but required greater computational resources.

The authors concluded that CNN-based models offer an effective baseline for ISL recognition but highlighted limitations such as reduced performance on dynamic gestures and real-time constraints. They recommended integrating temporal models to handle continuous sign sequences.

**Author 2: Reddy et al. (Hybrid CNN-LSTM Sign Recognition)**

Reddy and colleagues introduced a hybrid CNN-LSTM approach for dynamic gesture recognition. Their system captured both spatial features (via CNN) and temporal dependencies (via LSTM), making it suitable for continuous sign sentences. This improved recognition of motion-based ISL gestures.

Their experiments showed that incorporating optical flow and temporal smoothing further enhanced accuracy for fast or overlapping gestures. The hybrid model was evaluated on real-world datasets, demonstrating strong resilience to background noise and variations in user movement.

Despite improvements, the authors noted challenges such as high training time, dependency on high-quality video frames, and difficulty recognizing simultaneous hand motions. They proposed using transformer networks for improved parallel processing in future studies.

**Author 3: Kumar & Verma (Indian Speech Recognition Using Deep Neural Networks)**

Kumar and Verma investigated deep neural network-based speech recognition tailored for Indian languages and accents. Their system trained on large speech corpora, highlighting the benefits of deep learning over traditional HMM-based acoustic models. They achieved notable accuracy improvements in multi-accent scenarios. The authors implemented feature extraction using MFCC, filter banks, and spectrogram transformations. Using a DNN-HMM hybrid model, they improved recognition accuracy in noisy and reverberant environments. Their system balanced speed and accuracy for real-time applications. They concluded that Indian speech recognition can benefit from transformer-based architectures and larger contextual models. They emphasized the need for multilingual datasets to support India's diverse linguistic landscape.

**Author 4: Singh et al. (Transformer-Based Speech-to-Text Models)**

Singh and team worked on developing transformer-based speech-to-text systems with an emphasis on cross-lingual translation. Their model demonstrated strong performance due to the self-attention mechanism, which captures long-range dependencies more effectively than RNNs. Their research showed

that transformers outperform traditional models in both speed and accuracy, especially for long and complex sentences. They also demonstrated that pre-trained models such as Wav2Vec-2.0 significantly reduce training time and improve generalization. The authors noted that transformer models require substantial computational resources but offer state-of-the-art results. They recommended pruning, quantization, and model distillation for lightweight deployment.

**Author 5: Das & Nair (Assistive Communication Devices for the Deaf Community)**

Das and Nair reviewed various assistive communication tools designed for individuals with hearing impairments. These included glove-based systems, mobile applications, and wearable sensors. They analyzed limitations such as user discomfort and low accuracy in natural gesture recognition. Their study highlighted the shift toward camera-based systems using deep learning, which eliminate the need for wearable hardware. They identified CNN and RNN architectures as the most promising approaches for accurate gesture interpretation and real-time processing. The authors concluded that combining gesture recognition with speech processing could create holistic communication solutions. They emphasized the need for user-friendly interfaces and mobile-friendly deployment.

## III. EXISTING SYSTEM

Existing systems for sign language recognition typically focus on static gesture identification or dynamic gesture detection using glove sensors. Many applications rely on manual feature extraction and lack scalability across different lighting conditions and backgrounds. Similarly, existing speech recognition systems struggle with Indian accents and noise-rich environments. Currently available tools do not provide a unified interface for both ISL understanding and speech-to-text translation, limiting their usability for real-world communication between deaf and hearing individuals.

## IV. PROPOSED SYSTEM

The proposed system integrates Indian Sign Language recognition and speech-to-text translation into one unified, deep learning-driven framework. Using CNN-based gesture feature extraction and LSTM or transformer-based temporal modeling, the system recognizes both static and dynamic ISL gestures. Simultaneously, the speech recognition module uses transformer-based models such as Wav2Vec-2.0 to interpret spoken language with high accuracy. The system is designed for real-time operation, low latency, and user-friendly interaction, enabling seamless two-way communication.

## V. SYSTEM ARCHITECTURE

The architecture consists of two major modules—ISL recognition and speech recognition—connected to a central communication interface. The ISL module captures hand gesture frames using a camera, followed by preprocessing to remove noise and extract regions of interest. A CNN extracts spatial features, while an LSTM or transformer analyzes temporal motion for dynamic gestures. The recognized sign is converted into text for display. The speech recognition module takes audio input, applies noise filtering, and extracts acoustic features using spectrogram representations.

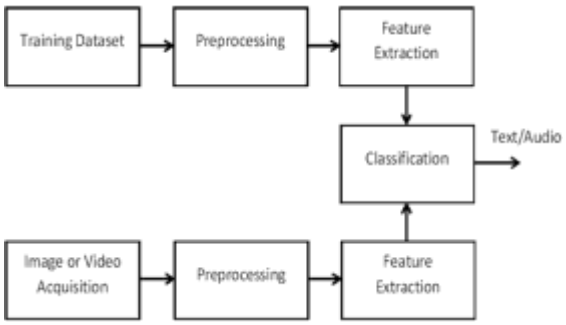


**Fig.5.1: System architecture**

A transformer-based speech-to-text model converts the audio into text with high accuracy. Both modules pass their outputs to a central interface that synchronizes communication and displays the interpreted text to users. The system supports real-time operation and can be deployed on mobile or embedded platforms.

## VI. IMPLEMENTATION

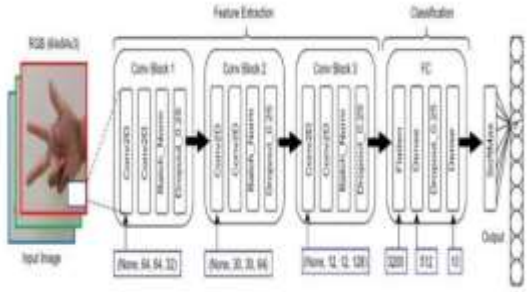

**Fig.6.1: CNN structure for hand gesture model**

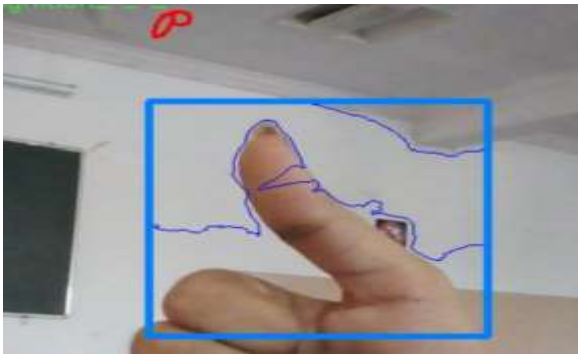

**Fig.6.2: Hand gesture input**


**Fig.6.3: Hand gesture output**

## VII. CONCLUSION

The integrated ISL and speech recognition system offers a powerful alternative to traditional assistive communication tools. By leveraging deep learning in both vision and speech domains, the system provides high accuracy, robustness, and real-time performance. It simplifies communication between deaf and hearing individuals by enabling two-way interaction without requiring any external hardware like gloves or sensors. The work demonstrates the potential of unified AI models to address accessibility challenges faced by the hearing-impaired community.

## VIII. FUTURE SCOPE

The future scope of the Indian Sign Language (ISL) and speech recognition system presents numerous opportunities for technological advancement, societal impact, and large-scale deployment. One major direction includes expanding the system to recognize a full spectrum of ISL components—such as hand

shapes, motion trajectories, facial expressions, mouth patterns, and upper-body gestures—allowing the model to interpret complex grammatical structures and convey complete contextual meaning rather than isolated words. Future research can integrate advanced transformer-based architectures, such as Vision Transformers (ViT) and Multimodal BERT models, to improve recognition accuracy, robustness, and generalization across diverse environments. Additionally, building larger, high-quality datasets representing different skin tones, lighting conditions, clothing contrasts, and regional ISL variations will help eliminate biases and enhance inclusivity. Real-time model optimization techniques, including edge AI deployment, model compression, and GPU–TPU acceleration, can be explored to enable seamless performance on smartphones, AR/VR devices, and wearable technology, making the system accessible to the broader population.

Another promising extension is the integration of multilingual speech recognition and translation capabilities, enabling the system to convert ISL into various Indian languages such as Hindi, Telugu, Tamil, Bengali, or Marathi, thereby increasing its utility across diverse linguistic communities. The platform could also incorporate emotion recognition, sign-to-avatar visualization, and context-aware dialogue generation to create a more natural and human-like communication experience. In the domain of accessibility, the system can be linked with educational platforms, virtual classrooms, healthcare assistance tools, and government service kiosks to support individuals with hearing or speech impairments. Furthermore, cloud-based datasets, federated learning, and secure model training can be implemented to protect user privacy while continuously improving recognition accuracy. In the long term, the system could evolve into a national-level communication infrastructure that bridges communication gaps in schools, workplaces, hospitals, public transportation systems, and emergency response services. With continued advancements in AI, sensor technology, and human–computer interaction, the proposed ISL and speech recognition framework has the potential to transform accessibility, empower millions with communication disabilities, and contribute significantly to an inclusive digital India.

## IX. REFERENCES

[1] K. M. J. R. A. &. R. I. Tiku, "Real-time Conversion of Sign Language to Text and Speech," in Tiku, K., Maloo, J., Ramesh, A., & R, I. (2020). Real-time Conversion of Sign Language to Text and Speech. 2020 Second International CSecond International Conference on Inventive Research in Computing Applications (ICIRCA), 2020.

[2] S. Y. M. M. K. S. V. S. &. S. S. Heera, "Talking Hands – An Indian Sign Language to Speech Translating Gloves," in International Conference on Innovative Mechanisms for Industry Applications (ICIMIA 2017), 2017.

[3] Hunter Phillips, Steven Lasch & Mahesh Maddumala , “American Sign Language Translation Using Transfer Learning”.

[4] Todupunuri, A. (2025). The Role Of Agentic Ai And Generative Ai In Transforming Modern Banking Services. American Journal of AI Cyber Computing Management, 5(3), 85-93.

[4]M. Rajmohan, C. Srinivasan, Orsu Ranga Babu, Subbiah Murugan, Badam Sai Kumar Reddy “Efficient Indian Sign Language Interpreter For Hearing Impaired”.

[5]Mahmudul Haque, Syma Afsha, Tareque Bashar Ovi, Hussain Nyeem, “Improving Automatic Sign Language Translation with Image Binarisation and Deep Learning”

[6] G. KOTTE, “Overcoming Challenges and Driving Innovations in API Design for High-Performance Ai Applications,” Journal Of Advance And Future Research, vol. 3, no. 4, 2025, doi: 10.56975/jaafr.v3i4.500282.

[6] K.Bhanu Prathap, G.Divya Swaroop, B.Praveen Kumar, Vipin Kamble, Mayur Parate, "ISLR: Indian Sign Language Recognition".
[7] Todupunuri, A. (2022). Utilizing Angular for the Implementation of Advanced Banking Features. Available at SSRN 5283395.
[7] Pavleen Kaur, Payel Ganguly, Saumya Verma, Neha Bansal, "Bridging the Communication Gap: With Real Time Sign Language Translation".
[8] G. Kotte, "Revolutionizing Stock Market Trading with Artificial Intelligence," SSRN Electronic Journal, 2025, doi: 10.2139/ssrn.5283647.
[8]Hao Zhou, Wengang Zhou, Weizhen Qi, Junfu Pu, Houqiang Li, "Improving Sign Language Translation with Monolingual Data by Sign Back-Translation".
[9] S. T. Reddy Kandula, "Comparison and Performance Assessment of Intelligent ML Models for Forecasting Cardiovascular Disease Risks in Healthcare," 2025 International Conference on Sensors and Related Networks (SENNET) Special Focus on Digital Healthcare(64220), pp. 1–6, Jul. 2025, doi: 10.1109/sennet64220.2025.11136005.
[9] Wanbo Li, Hang Pu, Ruijuan Wang, "Sign Language Recognition Based on Computer Vision".
[10] S. T. R. Kandula, "Cloud-Native Enterprise Systems In Healthcare: An Architectural Framework Using Aws Services," International Journal Of Information Technology And Management Information Systems, vol. 16, no. 2, pp. 1644–1661, Mar. 2025, doi: https://doi.org/10.34218/ijitmis_16_02_103
[10]Neeraj Kumar Pandey, Aakanchha Dwivedi, Mukul Sharma, Arpit Bansal, Amit Kumar Mishra, "An Improved Sign Language Translation approach using KNN in Deep Learning Environment".
[11]. R Vijaya Prakash, Akshay R, A Ashwitha Reddy, R Harshitha, K Himansee, S.K Abdul Sattar, "Sign Language Recognition Using CNN".
[12]. Sakshi Sharma, Sukhwinder Singh, "Vision-based sign language recognition system: A Comprehensive Review