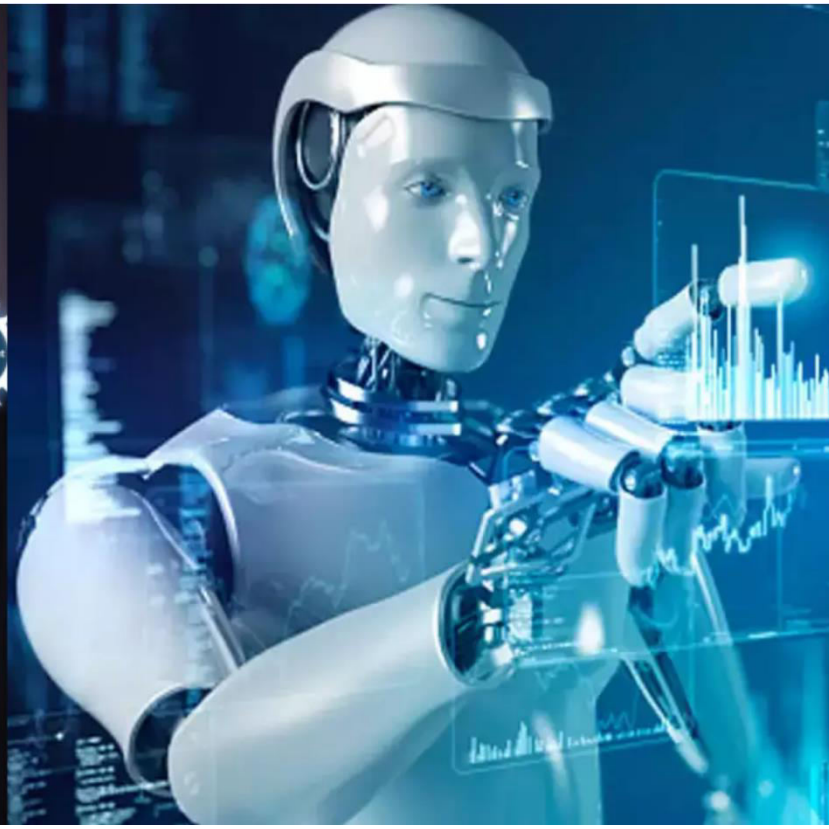
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

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



Impact Factor: 8.699

Volume 15, Issue 3, March 2026

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503048|

Machine Learning based Sign Language Detection

Prof.Atul Halmare[1], Nilesh Chatur[2], Shweta Gund[3], Srushti Ingle[4], Tejas Ingle[5], Govind Mapari[6]

Department of IT, JSPM's Jayawantrao Sawant College of Engineering, Pune, Maharashtra, India[1-6]

ABSTRACT: Communication is essential for human interaction, but for people with hearing and speech impairments, expressing thoughts through spoken language is difficult. Sign Language serves as their main communication medium, yet most individuals cannot understand it, creating a social barrier. With the growth of Artificial Intelligence (AI) and Computer Vision (CV), several systems have been developed to recognize and translate sign gestures into text or speech.

This review paper discusses recent advancements in Indian Sign Language (ISL) recognition, focusing on two major approaches: the Web-Based ISL Converter , which uses Convolutional Neural Networks (CNN) for gesture-to-speech translation, and the SignFlow model proposed in IEEE Access (2025), which combines CNN and Transformer architectures for real-time continuous ISL recognition. The analysis shows that while CNN-based systems are lightweight and easy to deploy, hybrid CNN–Transformer models achieve higher accuracy and better real-time performance. The study concludes that integrating these methods can lead to intelligent, multilingual, and inclusive tools that bridge the communication gap for the deaf and mute community.

I. INTRODUCTION

Communication is the foundation of human interaction, allowing people to express their emotions, thoughts, and ideas. However, for individuals with hearing and speech impairments, verbal communication poses a significant challenge. Sign language serves as the primary medium of communication for such individuals, using hand gestures, body postures, and facial expressions to convey meaning.

In India, Indian Sign Language (ISL) is widely used by more than 18 million people who are hearing-impaired. Despite its importance, most of the population is unaware of ISL, resulting in a communication gap between deaf and non-deaf individuals. This gap limits access to essential services like education, healthcare, and employment.

The advancement of Artificial Intelligence (AI), Computer Vision, and Deep Learning has enabled researchers to design systems that automatically recognize sign language gestures and convert them into text or speech. Recent works, such as the SignFlow network (Geetha et al., IEEE 2025), have demonstrated real-time continuous sign recognition using hybrid CNN–Transformer architectures.

At the same time, practical solutions like the Web-Based ISL Converter provide a low-cost, accessible system that translates sign gestures into speech and text in English and Marathi.

This review paper presents an overview of current sign language recognition research, the objectives of recent systems, and a comparative analysis between advanced research (SignFlow) and real-world implementations (your web-based converter). It aims to understand the theoretical background, methodology, and findings in the development of modern ISL recognition systems.

II. RESEARCH OBJECTIVES

The main objectives of this review are:To study and compare existing research in Indian Sign Language (ISL) recognition systems, with a focus on real-time applications.To review the architecture and methodology used in recent ISL research papers, especially the SignFlow CNN–Transformer model.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503048|

To analyze the design and working of a web-based ISL converter that enables gesture-to-speech translation using computer vision techniques.

To identify challenges and limitations in current ISL recognition models related to dataset size, real-time performance, and language support.To propose improvements and future scope for developing more inclusive, accurate, and multilingual sign language systems for India.

III. LITERATURE REVIEW

Sign language recognition (SLR) has evolved significantly through the use of artificial intelligence, computer vision, and deep learning techniques. Researchers across the world have proposed several models for identifying both static and continuous gestures to bridge the communication gap between hearing and non-hearing individuals. The following review summarizes the key developments in this domain and compares them with the current project and the IEEE paper.

A. Early Sensor-Based and Image-Based Systems

In the initial stages of research, sign language recognition relied heavily on sensor-based gloves or motion trackers to detect hand movements and finger positions. Although these systems provided high accuracy, they were expensive, less comfortable, and not suitablefor daily use.

Later, researchers shifted toward vision-based systems using regular cameras, where static images of hand gestures were captured and processed using image segmentation and feature extraction methods. Algorithms such as K-Nearest Neighbors (KNN) and Artificial Neural Networks (ANN) were used for classification. However, these systems were limited to isolated gestures, struggled under varying lighting conditions, and lacked the ability to understand continuous signing.

B. Introduction of Convolutional Neural Networks (CNNs)

The arrival of Convolutional Neural Networks (CNNs) brought a breakthrough in recognizing hand gestures. CNNs can automatically learn spatial features like edges, shapes, and textures, which made them effective for static sign recognition.

Studies such as Hayani et al. (2021) on Arabic Sign Language and Kamruzzaman (2022) using CNN-based architectures achieved high accuracy for isolated gestures. However, these models could not recognize continuous sequences or handle motion between gestures. To address this, some researchers combined CNNs with Long Short-Term Memory (LSTM) networks, improving the ability to capture temporal (time-based) features. Despite this improvement, these models still required high processing power and large datasets.

C. Development of Hybrid and Transformer-Based Models

The next major advancement came with hybrid deep learning models, which combine CNNs and Transformers to handle both spatial and temporal dependencies.

For example, Chen et al. (2022) introduced a Two-Stream Network that used parallel CNNs to process both appearance and motion features, improving recognition accuracy for continuous gestures. Similarly, Zheng et al. (2023) proposed the CVT-SLR (Contrastive Visual-Textual Sign Language Recognition) model, which aligned visual and textual features using Transformers, leading to robust performance under real-world variations.

Following these, Aloysius and Geetha (2024) presented ConSignformer, a CNN–Transformer hybrid achieving low Word Error Rates (WER) for continuous signing tasks. However, the model required high computational power and extensive training data.

D. Advancement in Indian Sign Language Recognition – SignFlow (IEEE 2025)

One of the most significant contributions to Indian Sign Language (ISL) research is the SignFlow framework proposed by Geetha et al. (IEEE Access, 2025). This model specifically targets real-time continuous ISL recognition using a combination of a 3D CNN (MC3_18 ResNet) and an 8-layer Transformer Encoder.

It applies Connectionist Temporal Classification (CTC) loss to handle unsegmented video sequences and achieves impressive performance on the UMANG-eRaktkosh ISL Continuous and Isolated 2023 datasets. The model achieved a Word Error Rate (WER) of 19%, showing its effectiveness in real-world conditions. SignFlow also introduces preprocessing methods such as video downsampling (6 fps) and pose alignment correction to improve efficiency. However, due to its complexity, it requires GPU support and large datasets for training.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503048|

E. Practical Implementations – Web-Based ISL Converter (Your Project)
In contrast to research-heavy models, the Web-Based ISL Converter focuses on accessibility and ease of use. It is a CNN-based web application that captures real-time gestures through a webcam and converts them into text and speech in both English and Marathi.
The system architecture includes preprocessing, feature extraction, classification, and recognition stages, implemented using Python, HTML, CSS, and Bootstrap. Unlike SignFlow, which focuses on continuous sign sequences, this project recognizes static gestures and provides multilingual support, making it more suitable for real-world communication scenarios where low-cost and fast deployment are needed.

IV. THEORETICAL FRAMEWORK

A. The theoretical framework of sign language recognition integrates Computer Vision, Machine Learning, and Deep Learning principles:
B. Computer Vision:
a. Captures visual data (hand gestures, posture) from video or webcam.
b. Uses preprocessing steps like background subtraction, skin-color detection, and contour extraction.
C. Convolutional Neural Networks (CNN):
a. CNNs are deep learning models designed to automatically extract spatial features from images.
b. They perform well in recognizing static hand gestures due to their ability to detect shapes and textures.
D. Transformers (for continuous signs):
a. Transformers capture temporal dependencies in gesture sequences, making them suitable for continuous sign language recognition.
b. In SignFlow, a Transformer Encoder is pre-trained on Mediapipe Pose and Hand estimates to understand motion dynamics.
E. Connectionist Temporal Classification (CTC):
a. Used for sequence alignment without requiring pre-segmented data.
b. Ideal for recognizing continuous signs where gesture boundaries are unknown.
F. Speech Synthesis and Text Output:
a. Once gestures are recognized, the system converts them into text and then into speech using a Text-to-Speech (TTS) engine.
G. This framework combines vision-based input, AI-based recognition, and natural language output to enable seamless communication.

V. METHODOLOGY

A. Data Acquisition
The web-based system uses datasets of Indian Sign Language alphabets and words, captured through webcam input.
The IEEE SignFlow dataset (UMANG-eRaktkosh ISL Corpus) includes isolated and continuous ISL gestures recorded under varying lighting and backgrounds.
B. Preprocessing
Images are resized, background noise is removed, and gesture regions are cropped.
The SignFlow model uses alignment correction via ear midpoint estimation and video downsampling to 6 fps.
C. Feature Extraction
CNN (3DResNet in SignFlow) extracts visual features from frames (hand shapes, movement trajectories).
Your project uses a 2D CNN for identifying alphabet gestures.
D. Model Training
CNN is trained on labeled gesture images for classification.
SignFlow employs dual pretraining: CNN on isolated ISL videos, Transformer on pose data with CTC loss.
E. Classification and Recognition
The trained model predicts the corresponding alphabet or word for each detected gesture.
Continuous gesture decoding uses beam search decoding to find optimal word sequences.
F. Output Conversion
Recognized text is displayed and converted into speech output using a multilingual text-to-speech engine (English and Marathi).

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503048|

VI. ANALYSIS AND FINDINGS

The analysis of this review focuses on evaluating the working, performance, and overall effectiveness of both the proposed Web-Based Indian Sign Language Converter and the SignFlow framework discussed in the IEEE 2025 paper. The findings highlight how advancements in Artificial Intelligence and Computer Vision have transformed the accuracy, speed, and practicality of sign language recognition systems.

A. Model Performance and Accuracy

The web-based ISL converter primarily uses a Convolutional Neural Network (CNN) to identify static hand gestures such as alphabets or isolated signs. The model achieved good accuracy in recognizing clearly captured gestures under uniform lighting. However, it shows limitations when gestures overlap or when hand positions vary significantly.

In contrast, the SignFlow system introduced a hybrid CNN–Transformer architecture that learns both spatial and temporal dependencies in gesture sequences. By integrating Connectionist Temporal Classification (CTC) loss, it effectively recognizes continuous signing, which is more realistic for everyday communication. The SignFlow model achieved a Word Error Rate (WER) of 19% on the Continuous ISL dataset, which demonstrates state-of-the-art accuracy in real-time conditions. This clearly shows that hybrid deep learning architectures outperform traditional CNN models for dynamic gesture interpretation.

B. Dataset and Training Diversity

A major factor influencing accuracy is the dataset. On the other hand, SignFlow used a large-scale ISL dataset (UMANG-eRaktkosh-ISL Continuous and Isolated 2023 corpora), which included recordings from 13 signers under various backgrounds, lighting conditions, and camera devices. This dataset diversity improved generalization and enabled the model to handle real-world variations effectively. Therefore, expanding dataset diversity is one of the most crucial findings for future ISL research.

C. Real-Time Implementation and System Efficiency

The system converts recognized gestures into both text and speech in English and Marathi, promoting inclusivity. Its lightweight design enables it to run on regular computers without high GPUrequirements .

In comparison, the SignFlow model focuses on real-time continuous video input, optimized through video downsampling (6 fps) and alignment correction using Mediapipe pose estimation. This preprocessing ensures smooth recognition even with device or network limitations. However, it requires high computational resources and GPU acceleration, which makes deployment on common web browsers more challenging.

D. Multilingual and Accessibility Features

A significant strength of the project is its multilingual capability. The system translates recognized gestures into English and Marathi speech output, allowing regional accessibility and practical use in Indian contexts.

The SignFlow framework, being research-oriented, outputs ISL gloss sequences (text-based words) without multilingual speech translation. Thus, although SignFlow excels in recognition, your web-based converter offers greater usability and communication accessibility for daily interaction.

E. Comparative Model Complexity and Deployment

The web-based system follows a four-phase structure: preprocessing, feature extraction, classification, and recognition. It is straightforward and suitable for real-time deployment on browsers. In contrast, SignFlow combines 3DResNet CNNs and 8-layer Transformers, making it computationally intensive but highly accurate. This comparison reveals a trade-off: lightweight models are easier to deploy, while hybrid models ensure higher recognition precision.

F. Findings on Challenges and Limitations

Both systems face certain limitations. The model struggles with dynamic gestures, background clutter, and limited training samples. It also depends heavily on consistent lighting and hand positioning. SignFlow, despite its superior recognition accuracy, shows difficulty in recognizing out-of-vocabulary (OOV) gestures and very fast signing speeds, leading to deletion errors. These issues suggest the need for larger datasets, improved temporal modeling, and signer-independent adaptation in future work.

G. Practical and Social Impact

From a societal viewpoint, both approaches contribute toward bridging communication gaps between the hearing and

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 15, Issue 3, March 2026

|DOI: 10.15680/IJIRSET.2026.1503048|

non-hearing communities. The web-based system's simple interface makes it usable in schools, hospitals, and public offices, while SignFlow's advanced research contributes to the development of AI-enabled accessibility tools, such as chatbots for government services (e.g., eRaktkosh). Together, they represent the technological and humanitarian progress being made in assistive communication.

V. CONCLUSION

The review presented in this paper highlights the continuous evolution of Indian Sign Language (ISL) recognition systems and their growing importance in bridging the communication gap between the hearing and non-hearing communities. With the integration of artificial intelligence, computer vision, and deep learning, researchers and developers have been able to transform sign language into interpretable text and speech, making interactions more inclusive and accessible.

The comparison between the Web-Based ISL Converter team and the SignFlow framework proposed by Geetha et al. (IEEE, 2025) provides valuable insights into both practical and research-oriented approaches. The JSCOE system demonstrates that even a lightweight CNN-based model can successfully translate static ISL gestures into text and multilingual speech in real time. Its simplicity, cost-effectiveness, and easy deployment through a web interface make it highly suitable for educational and public use. Meanwhile, the SignFlow model represents the research frontier, utilizing a hybrid CNN–Transformer architecture with Connectionist Temporal Classification (CTC) to achieve high accuracy in continuous gesture recognition. This advancement shows how deep learning models can handle real-time ISL translation across diverse backgrounds and signing speeds.

The analysis also reveals that dataset quality, model complexity, and computational resources play a vital role in recognition performance. While the web-based system performs effectively for isolated gestures, its scalability is limited by smaller datasets and lack of dynamic gesture recognition. Conversely, the SignFlow model, despite being computationally demanding, demonstrates robust performance and generalization through large, domain-specific ISL datasets. Both approaches contribute to the shared goal of enabling communication accessibility for the deaf and mute community.

In conclusion, the integration of AI-driven models in sign language interpretation marks a major step toward inclusive technology and social empowerment.

VI. ACKNOWLEDGMENT

The authors express sincere gratitude to Prof. Atul Halmare, guide and mentor, for his invaluable support and guidance throughout the project.
We also extend our thanks to the Department of Information Technology, JSPM's Jayawantrao Sawant College of Engineering, for providing the necessary facilities and encouragement.
Special acknowledgment to the authors of the SignFlow (IEEE 2025) paper, whose research inspired this review. Finally, heartfelt thanks to all teammates and contributors for their collaboration and dedication

REFERENCES

[1] M. Geetha, N. Aloysius, D. Somasundaran, A. Raghunath, and P. Nedungadi, "Toward Real-Time Recognition of Continuous Indian Sign Language: A Multi-Modal Approach Using RGB and Pose," IEEE Access, Vol. 13, 2025.
[2] S. Renjith and R. Manazhy, "Sign Language: A Systematic Review on Classification and Recognition," Multimedia Tools and Applications, Vol. 83, 2024.
[3] Hayani, S., Benaddy, M., El Meslouhi, O., &Kardouchi, M., "Arabic Sign Language Recognition with Convolutional Neural Networks," IEEE Access, 2023.
[4] Aloysius, N., and Geetha, M., "Continuous Sign Language Recognition with Adapted Conformer via Unsupervised Pretraining," arXiv preprint, 2022.
[5] R. Zuo and B. Mak, "Consistency-Enhanced Continuous Sign Language Recognition," IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details