iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: XI Month of publication: November 2025

DOI: https://doi.org/10.22214/ijraset.2025.75069

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

Real-Time Sign Language Recognition and Translation: A Survey of Deep Learning Techniques

B. N. Madhukar[1], Nishanth P[2], Naveen Kumar V[3], Revansidda[4], Sakshi Yeli[5]
[1]Assistant Professor, Electronics and Communication Dept., AMCEC, Karnataka, India
[2, 3, 4, 5]Student, Electronics and Communication Dept., AMCEC, Karnataka, India

***Abstract:** Closing the communication gap between the Deaf and Hard-of-Hearing (DHH) community and hearing people has remained a key area of research in the recent past. This survey paper discusses the most current deep learning methods applied in Sign Language Recognition (SLR) and Sign Language Translation (SLT) systems. The paper examines the transition from the traditional vision-based methods to more recent neural models, such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Transformer models such as Sign Language Transformers (SLT). Hybrid models that integrate CNN and LSTM have scored remarkable accuracy rates of more than 90% in static as well as dynamic gestures. The latest YOLO and MediaPipe frameworks enable real-time detection of hand gestures through simple webcams. The paper also compares various datasets, performance metrics, and preprocessing techniques utilized for several regional sign languages, such as Indian, Arabic, and American Sign Languages. Challenges still remain in continuous sign recognition, signer variation, and dataset variation despite the major advancements. This work recommends directions for the development of fully end-to-end, real-time, and multilingual sign language interpretation systems toward inclusive human-computer interaction.*

***Keywords:** Sign Language Recognition, Deep Learning, CNN, LSTM, Transformer, MediaPipe, YOLO, Real-Time Systems.*

I. INTRODUCTION

Sign language is a crucial form of communication for the Deaf and Hard-of-Hearing (DHH) community, conveying meaning through hand signs, facial expressions, and body language. But sign and non-signer communication barriers restrict accessibility in education, healthcare, and public services. Sign Language Recognition (SLR) and Sign Language Translation (SLT) seek to overcome this barrier by translating signs into text or speech through artificial intelligence (AI). The convergence of deep learning, computer vision, and natural language processing (NLP) has greatly enhanced the accuracy of gesture recognition and real-time translation. Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer-based models have shown great potential for learning spatial and temporal features. This paper gives an extensive overview of recent work on deep learning-based SLR and SLT, including leading models, datasets, metrics, and comparisons of performance to establish present trends and future directions.

II. DATA ACQUISITION AND PREPROCESSING

The basis of any sign language recognition system rests upon the efficient collection and preprocessing of data. Visual-based data collection is usually done through the use of cameras, webcams, or sensors like Microsoft Kinect. Libraries like MediaPipe are commonly used to get 21 key points from the hands and 33 pose markers from the body, allowing for efficient hand and pose tracking. Preprocessing methods involve frame extraction, ROI segmentation, normalization, and background subtraction so that the analysis can be performed on only those gesture features that are relevant. In real-time systems, light feature extraction mechanisms and effective data augmentation are essential to keep recognition accuracy invariant to changing lighting and signer conditions. Standard databases like ASL Alphabet, RWTH-PHOENIX-2014T, and regional sign language databases created specifically for an application are commonly utilized to compare performance.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XI Nov 2025- Available at www.ijraset.com

III. LITERATURE SURVEY

Using deep learning models, SLR and SLT have advanced rapidly in recent years. GestureNet, a hybrid CNN-based system with 92% accuracy in real-time Indian Sign Language (ISL) recognition, was proposed by Narayan et al. [1]. For 3D skeletal gesture recognition, Bayegizova et al. [2] used a CNN–LSTM model, demonstrating strong performance across signers. Outperforming previous RNN-based models, Camgoz et al. [3] introduced Sign Language Transformers (SLT), an end-to-end system that combines recognition and translation without gloss-level supervision. A systematic review of more than 85 studies from 2020 to 2024 was carried out by Rao et al. [4], who emphasized Transformer architectures as the new standard in CSLR. [5] Noor et al.created a hybrid CNN–LSTM for Arabic Sign Language that achieved 94.4% accuracy, while Alsharif et al. [6] used MediaPipe keypoint tracking and YOLOv11 to detect ASL in real time with a mean average precision of 98.2%. With remarkable accuracy, Khan et al. [7] created a transfer learning model for ASL alphabets using ResNet50 and MobileNetV2. Yilmaz et al. [8] achieved high detection precision real-time gesture recognition by implementing YOLOv8 for Turkish Sign Language. Singh et al. [9] used a multi-modal Transformer framework for Sign-to-Text translation, combining computer vision and natural language processing. Lastly, a comparative deep learning review by Patel et al. [10] demonstrated the superiority of Transformer-based and hybrid architectures for both isolated and continuous gestures.

Ref.	Author(s) & Year	Model / Technique	Dataset Used	Accuracy / Metric	Key Findings / Contribution
[1]	A. Narayan et al., 2024	Hybrid CNN (GestureNet)	Custom ISL	92%	Compact hybrid model tailored for ISL gesture recognition.
[2]	A. Bayegizova et al., 2022	CNN–LSTM	Kinect / MediaPipe	90%	3D skeletal gesture modeling using combined spatial-temporal learning.
[3]	N. C. Camgoz et al., 2020	Sign Language Transformer (SLT)	PHOENIX-2014T	BLEU-4: 20.17	End-to-end Transformer for simultaneous recognition and translation.
[4]	Y. S. N. Rao et al., 2024	Systematic Review	Multiple Datasets	-	Surveyed over 85 models highlighting Transformer dominance in CSLR.
[5]	T. H. Noor et al., 2024	Hybrid CNN–LSTM	Custom ArSL	94.4%	Achieved high-accuracy recognition for Arabic Sign Language.
[6]	B. Alsharif et al., 2025	YOLOv11 + MediaPipe	Custom ASL	mAP@0.5 = 98.2%	Enhanced real-time ASL detection using keypoint-based YOLO.
[7]	M. Khan et al., 2023	ResNet50 / MobileNetV2	ASL Dataset	97%	Transfer learning approach for ASL alphabet recognition.
[8]	A. Yilmaz et al., 2024	YOLOv8 Framework	Turkish SL Dataset	96%	High-speed, precise sign recognition using YOLOv8.
[9]	P. Singh et al., 2023	Multi-Modal Transformer	Custom Dataset	BLEU-4: 21.3	Sign-to-text translation integrating visual and linguistic data.
[10]	R. Patel et al., 2024	Comparative Review	Multiple	-	Comprehensive analysis confirming Transformer and hybrid model superiority.

Table 1: Comparative Performance of SLR and SLT Models

IV. ADVANCEMENTS AND EVALUATION

CNN-only models have given way to hybrid CNN–LSTM and Transformer-based architectures that manage temporal and spatial dependencies concurrently, as evidenced by recent advancements in sign language processing. These days, real-time recognition systems use MediaPipe for precise keypoint extraction and YOLO for detection. While new regional datasets for Arabic and Indian sign languages are starting to appear, the PHOENIX-2014T dataset is still a significant benchmark for SLT. Metrics like accuracy, Word Error Rate (WER), and BLEU score are used to evaluate performance. Since Transformers were introduced, BLEU scores have increased from 9.58 to over 21, demonstrating superior translation ability. Furthermore, real-time inference with consumer-grade GPUs or embedded systems is now possible thanks to improvements in hardware and optimization.

V. CONCLUSION

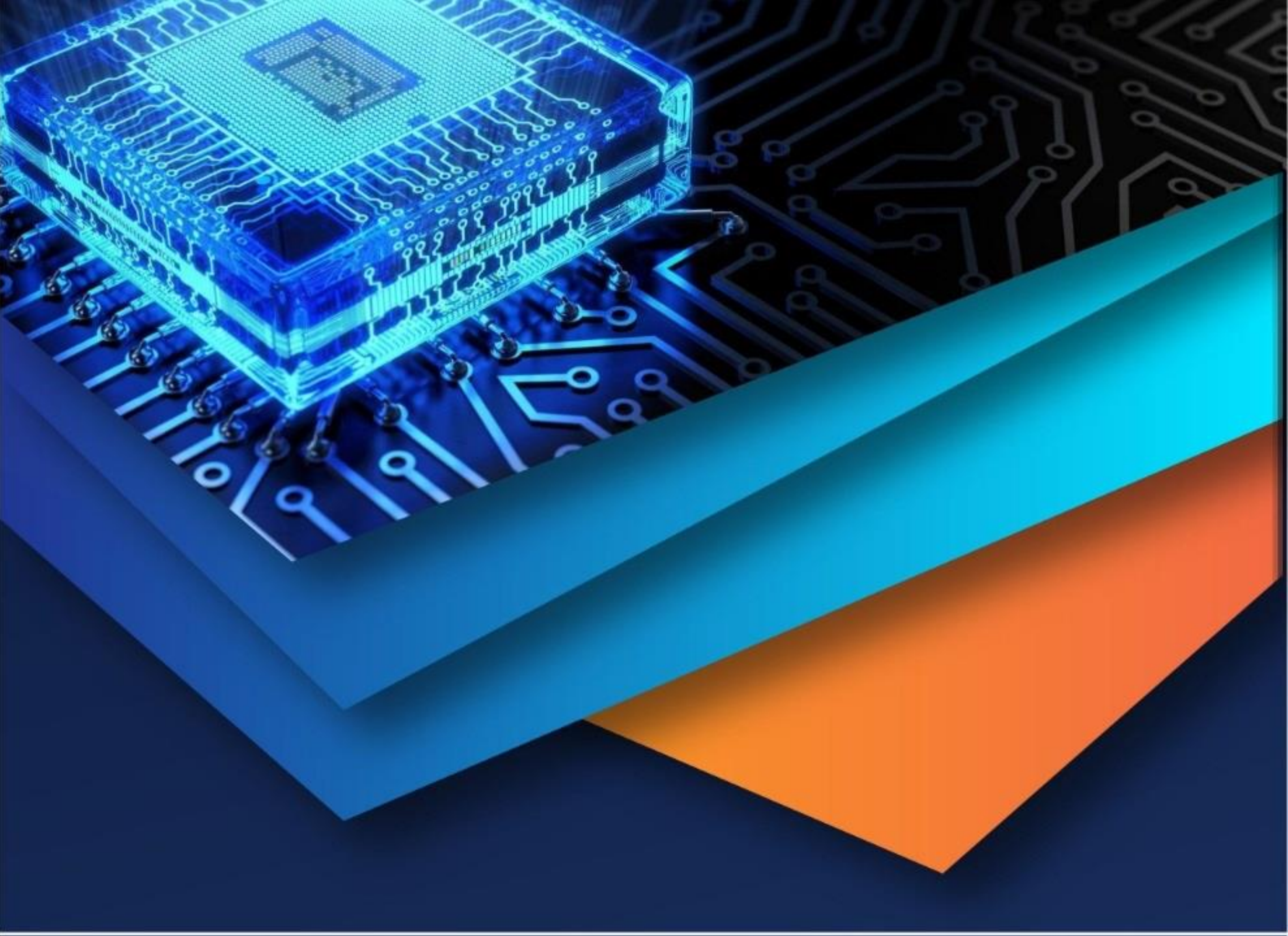
Sign language translation and recognition have been transformed by deep learning, greatly increasing accessibility for the DHH community. Models, ranging from CNN-based classifiers to end-to-end Transformer networks, have demonstrated exceptional accuracy in multiple languages, surpassing 90%. Notwithstanding these successes, issues like signer variability, illumination, and a small number of diverse datasets still exist. For more realistic translation, future studies should concentrate on multilingual, real-time SLT systems that incorporate lip movement, facial expression, and gesture. AI-powered sign language systems have the potential to develop into effective instruments for inclusive communication with further development.

VI. ACKNOWLEDGEMENT

The authors express their gratitude to all researchers and open-source contributors whose work has advanced deep learning applications for sign language interpretation.

REFERENCES

[1] Narayan et al., GestureNet: Real-Time Sign Language Recognition Using a Hybrid Neural Network, IJFMR, 2024.
[2] A. Bayegizova et al., Effectiveness of the Use of Algorithms and Methods of Artificial Technologies for Sign Language Recognition, EEJET, 2022.
[3] N. C. Camgoz et al., Sign Language Transformers, arXiv:2003.13830, 2020.
[4] Y. S. N. Rao et al., Dynamic Sign Language Recognition and Translation, JTAIT, 2024.
[5] T. H. Noor et al., Real-Time Arabic Sign Language Recognition, Sensors, 2024.
[6] B. Alsharif et al., Real-Time American Sign Language Interpretation, Sensors, 2025.
[7] M. Khan et al., Deep Learning Technology to Recognize ASL Alphabet, Sensors, 2023.
[8] A. Yilmaz et al., Real-Time Sign Language Recognition Based on YOLO Algorithm, NCA, 2024.
[9] P. Singh et al., Sign Language Recognition and Translation: A Multi-Modal Approach, ACL Anthology, 2023.
[10] R. Patel et al., Deep Learning for Sign Language Recognition: A Comparative Review, Paradigm, 2024.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY