Sign Language Recognition: A Comprehensive Review Of Methods, Datasets, And Challenges

A.BalaAyyappan[1], Dr.T.Gobinath[2], A.Sivaramakrishnan[3], Dr.M.Kumar[4], P.Srimathi[5], T.Selvathiyaneshwaran[6]
[1, 5, 6]Dept of Artificial Intelligence and Data Science
[2, 3]Dept of Computer Science and Engineering
[4]Dept of Electronics and Communication Engineering
[1, 2, 3, 4, 5, 6] Chettinad College of Engineering and Technology, Karur, India

Abstract- *Sign Language Recognition (SLR) is an active research area that involves the automatic translation of sign languages into text or speech using computational techniques, thereby facilitating communication for deaf and hard-of-hearing individuals. Over the past two decades, SLR has evolved from traditional handcrafted feature-based approaches to modern deep learning-driven methods due to significant advances in computer vision. This survey reviews approximately 15–25 representative studies and categorizes existing SLR approaches into isolated and continuous recognition tasks. Widely used datasets, feature extraction techniques, learning models, and evaluation metrics are discussed. Classical methods such as Hidden Markov Models and Support Vector Machines are reviewed alongside deep neural architectures including convolutional neural networks, recurrent neural networks, and transformer-based models. Key challenges include data sparsity, signer variability, and difficulties in modeling non-manual cues. Finally, promising future research directions such as multimodal learning and low-resource sign language recognition are highlighted.*

Keywords- Sign Language Recognition, Indian Sign Language, Deep Learning, CNN, ISLR, CSLR

I. INTRODUCTION

Sign language is a natural visible language used by the deaf and hard-of-hearing community, communicating by hand gestures, facial expressions, and body movements. Indian Sign Language (ISL) is one of the most widely used sign languages in India; communication between ISL users and hearing people, on account of unsuccessful implementation of automated interpretation systems, is still difficult. Sign Language Recognition allows the computers to automatically trace the interpretations of the signs. SLR early research focus was given to isolated sign language recognition, in which individual signs are performed independently. Most such sys- tems are based on handcrafted features such as hand shape descriptors, motion trajectories and optical flow are combined with traditional machine learning models like Support Vector Machines(SVM) and Hidden Markov Models (HMM). While these methods realized reasonable performance, they were highly sensitive to variations in illumination, background clutter, and signer specific style. Recent SLR research, with developments in computer vision and deep learning, has taken a turn toward data-driven approaches: Spatial features in sign images and videos have been extracted using CNNs, while RNNs and 3D CNNs have been adopted to model temporal dynamics in sign sequences. More recently, attention mechanisms and transformer-based architectures have shown improved performance by effectively capturing long-range temporal dependencies. Nevertheless, some technologies have been developed in the field of ISL. Comparing the number of studies carried out in ISL with those of other widely studied sign languages like American Sign Language or Chinese Sign Language, it would be found that there is less work in ISL. This survey provides an overview of some relevant studies of Sign Language Recognition, especially Isolated Sign Language Recognition, in ISL. The objective of this survey is to understand the issues in ISL-based Sign Language Recognition.

II. PROBLEM FORMULATION AND TASK DEFINITION

Here, we can understand the various sign language methods like ISLR (Isolated sign language recognition) and CSLR (Continuous sign language recognition).

A. *Isolated Sign Language Recognition(ISLR)*

Alsharif et al. (2024) achieve SOTA 98% mAP on isolated ASL alphabet recognition using YOLOv8+MediaPipe land- marks on 29,820 merged images. Their two-stage pipeline first localizes hands via transfer learning from COCO, then annotates 21-point hand landmarks for precise training. Out- performs YOLOv9 (97.84% mAP) with 11.8 FPS on consumer CPU, though confusions persist between similar gestures (M/N, O/P). Represents performance ceiling for single-frame alphabet detection but doesn't address ISL's 4,800+ lexical items or regional dialects [1].

B. *Continuous Sign Language Recognition(CSLR)*

YOLOv4 for detection + SVM+MediaPipe for landmark- based classification. An expert system dynamically selects higher-confidence predictions from both models during real- time webcam inference. Gramformer post-processes static gesture sequences into grammatically correct sentences (”Are you thirsty?”). Self-created 676-image dataset (augmented to 13,000) targets India's two-handed ISL vocabulary. Establishes first scalable continuous ISL recognition benchmark outper- forming prior YOLOv3 baselines [2].

III. DATASETS AND DATA MODALITIES

Most Sign Language Recognition (SLR) studies rely on *RGB video data*, as it effectively captures hand gestures and body movements using standard cameras. Recently, pose-based representations have also been explored to reduce back-ground noise and improve recognition performance. Table I summarizes commonly used datasets along with their sign languages, tasks, and data modalities.

TABLE I
Commonly Used Datasets in Sign Language Recognition

Dataset	Sign Language	Task	Modality	Reference
MS-ASL	American SL	ISLR	RGB	[3]
WLASL	American SL	ISLR	RGB	[4]
AUTSL	Turkish SL	ISLR	RGB	[5]
INCLUDE	Indian SL	ISLR	RGB	[6]
I-SIGN	Indian SL	ISLR	RGB	[7]

IV. FEATURE REPRESENTATION TECHNIQUES

A. *Handcrafted Feature-Based Approaches*

The propose HOG-9ULBP methodology for MSMA hand gesture recognition. SGM-Kmeans segmentation first separates gestures from complex backgrounds—SGM handles non-skin while K-means (K=3) clusters black/skin/skin-like pixels for skin-like interference. HOG extracts scale-invariant contour features via 9-bin gradient histograms (64×64 blocks, 8×8 cells). 9ULBP improves texture extraction by limiting uni- formity (U1) in binary patterns, selecting minimal LBP code per 1s-count class for rotation invariance. Feature fusion cascades normalized HOG+9ULBP descriptors, fed to RBF-SVM classifier [8].

B. *Deep Feature Learning Approaches*

CNNs serve as baseline classifiers and feature extractors in prior sign language recognition works referenced in the paper, such as CNN with LeNet-5 for feature extraction and classifi- cation of SL images.They enable speech-to-SL translation via gesture generation and direct ISL alphabet recognition, though limited by dataset imbalances and lack of temporal modeling [9].

V. RECOGNITION MODELS

A. *Classical Machine Learning Models*

The paper fuses skeleton-based handcrafted ML fea-tures—distances (190) between joints and angles (630) from direction vectors—with GoogleNet's 1024 pixel-based deep features for JSL recognition. Boruta with Random Forest selects optimal features (e.g., 777 from 1843), fed to kernel SVM (polynomial/RBF/sigmoid via one-vs-rest) for multi-class classification.Prior ML works include SVM on angles (63.6%), random forest on skeletons (70%), and decision trees with genetics (74%), all limited by single-feature reliance [10].

B. *Deep Learning Models*

DeepASLR introduces a novel American Sign Language Alphabet (ASLA) dataset featuring 104,000 images captured under variable lighting and distances to improve real-world robustness. The methodology employs a sequential Convo- lutional Neural Network (CNN) architecture with three con- volutional layers, ReLU activation, and MaxPooling for op- timized feature extraction. Evaluations demonstrate a high classification accuracy of 99.38% on the custom dataset, outperforming several existing state-of-the-art models in sign language recognition. The system also includes a real-time prediction interface utilizing OpenCV to translate static hand gestures into natural language [11] This paper presents a 3D- CNN architecture designed to recognize 20 dynamic gestures from Indian Sign Language (ISL) by capturing both spatial and temporal features from video sequences. The researchers developed a custom dataset of 2,400 videos across various lighting conditions and orientations to ensure environmental robustness. The proposed model achieved a training accuracy of 99.67% and a validation accuracy of 88%, outperforming existing state-of-the-art methods in ISL recognition. [12]

VI. EVALUATION PROTOCOLS

Application

ArSLR-ML is an interactive Python-based application that uses MediaPipe and multi-class SVM to recognize static Ara- bic Sign Language letters and numbers. The system captures gestures via a standard laptop camera and provides real- time translation into both text and Arabic audio output. It achieved high accuracy rates of 99.09% for alphabets and 98.08% for digits, with very low average response times. This cost-effective tool is specifically designed to support the education and social integration of hearing-impaired children [13]. This study presents a real-time assistive prototype fea- turing a sensory glove and an Android application to facilitate communication for the deaf and mute. By utilizing five flex sensors and an accelerometer, the system captures Arabic Sign Language gestures and transmits the data via Bluetooth to a mobile device. The application then converts these signals into both visual text and audible sound, providing a low- cost, portable solution to bridge the acoustic communication gap. This technology aims to improve social integration and

TABLE II

Comparative Analysis of Sign Language Recognition Methods

Author(s)	Year	Lang.	Dataset (Size)	Method / Architecture	Accuracy
Almjally & Almukadi [9]	2023	SL	Custom (15K images, 10 classes)	BF + ResNet-152 + Bi-LSTM + HHO	98.95%
Shin et al. [10]	2022	JSL	Lab dataset (7,380 images, 41 classes)	Skeleton features + GoogleNet + Boruta + SVM	98.53%
Alsharif et al. [1]	2024	ASL	Lexset + Kaggle + Roboflow (29,820 images)	YOLOv8 + MediaPipe	98% mAP
Sreemathy et al. [2]	2023	ISL	Self-created ISL (676, 13K augmented)	YOLOv4 + SVM + MediaPipe	98.8%
Kasapbas¸ı et al. [11]	2022	ASL	Custom ASLA + benchmarks	Sequential CNN (3 Conv layers)	99.48%
D. K. Singh [12]	2021	ISL	Dynamic ISL gestures (20 classes)	3D-CNN (Spatio-temporal)	88%

career opportunities for hearing-impaired individuals by en- abling easier interaction with the general public [14]. This research focuses on developing a Japanese Finger-spelled Sign Language (JFSL) recognition system to assist communication for hearing-impaired individuals. The application uses a depth sensor to precisely extract hand regions without requiring wearable landmarks like wristbands or electronic gloves. By combining Time-Series Curve (TSC) features with a deep neural network, the system translates complex gestures into natural language for real-time interaction. This approach offers a practical, non-invasive solution to enhance social integration and domestic accessibility for the deaf community [15].

VII. CHALLENGES AND OPEN ISSUES

Despite significant progress in Sign Language Recognition (SLR), several challenges and open issues remain, as identified in the surveyed literature. One of the primary challenges is the limited availability of large-scale and balanced datasets, particularly for low-resource sign languages such as Indian Sign Language. Many existing datasets contain a restricted number of signers and classes, resulting in data imbalance and biased model performance.

Another major challenge is the large variation in signing styles across individuals. Differences in hand shape, signing speed, regional variations, and personal expression signifi- cantly affect recognition accuracy. Models trained on limited signer diversity often fail to generalize well to unseen users.

Modeling non-manual cues, including facial expressions, head movements, and body posture, remains difficult. Most existing SLR systems primarily focus on hand gestures while neglecting non-manual components, which are linguistically important in sign languages.

Finally, poor generalization and robustness in real-world en- vironments remain open issues. Many approaches demonstrate high accuracy in controlled laboratory settings but degrade significantly under real-world conditions involving cluttered backgrounds, illumination changes, and varying camera view- points.

VIII. FUTURE RESEARCH DIRECTIONS

Future research in SLR should focus on the development of larger, more diverse, and well-annotated datasets, especially for underrepresented and low-resource sign languages. Incor- porating cross-dataset and cross-signer evaluation protocols can further enhance model robustness and generalization.

Another promising research direction is the design of mul- timodal SLR systems that integrate RGB video with skele- tal pose and facial landmark information. Such multimodal approaches can better capture both manual and non-manual components of sign language.

The development of lightweight and real-time SLR models is also essential for practical deployment on mobile and embedded platforms. Techniques such as model compression, knowledge distillation, and efficient network architectures can facilitate real-time performance.

Additionally, low-resource learning strategies, including transfer learning, self-supervised learning, and few-shot learn- ing, offer promising solutions for reducing reliance on large annotated datasets.

IX. CONCLUSION

This paper presented a simple and structured survey of 15– 25 representative studies on Sign Language Recognition. The survey highlighted the evolution of SLR systems from hand- crafted feature-based approaches to modern deep learning- based models. Although recent methods have achieved no- table performance improvements, challenges related to data availability, signer variability, non-manual cue modeling, and real-world deployment remain unresolved. Addressing these challenges is crucial for advancing practical and inclusive SLR systems, motivating continued research in this area.

REFERENCES

[1] B. Alsharif, E. Alalwany, and M. Ilyas, "Transfer learning with yolov8 for real-time recognition system of american sign language alphabet," *Franklin Open*, vol. 8, p. 100165, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2773186324000951

[2] R. Sreemathy, M. Turuk, S. Chaudhary, K. Lavate, A. Ushire, and S. Khurana, "Continuous word level sign language recognition using an expert system based on machine learning," *International Journal of Cognitive Computing in Engineering*, vol. 4, pp. 170–178, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666307423000177

[3] H. R. V. Joze and O. Koller, "MS-ASL: A large-scale data set and benchmark for understanding american sign language," *CoRR*, vol. abs/1812.01053, 2018. [Online]. Available: http://arxiv.org/abs/1812.01053

[4] D. Li, C. R. Opazo, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," *CoRR*, vol. abs/1910.11006, 2019. [Online]. Available: http://arxiv.org/abs/1910.11006

[5] O. M. Sincan and H. Y. Keles, "Autsl: A large scale multi- modal turkish sign language dataset and baseline methods," *IEEE Access*, vol. 8, pp. 181 340–181 355, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:220936609

[6] A. Sridhar, R. G. Ganesan, P. Kumar, and M. M. Khapra, "Include: A large scale dataset for indian sign language recognition," *Proceedings of the 28th ACM International Conference on Multimedia*, 2020. [Online]. Available: https://api.semanticscholar.org/CorpusID:222278681

[7] A. Joshi, R. Mohanty, M. Kanakanti, A. Mangla, S. Choudhary, M. Barbate, and A. Modi, "iSign: A benchmark for Indian Sign Language processing," in *Findings of the Association for Computational Linguistics: ACL 2024*, L.-W. Ku, A. Martins, and Srikumar, Eds. Bangkok, Thailand: Association for Computational Linguistics, Aug. 2024, pp. 10 827–10 844. [Online]. Available: https://aclanthology.org/2024.findings-acl.643/

[8] J. Li, C. Li, J. Han, Y. Shi, G. Bian, and S. Zhou, "Robust hand gesture recognition using hog-9ulbp features and svm model," *Electronics*, vol. 11, no. 7, 2022. [Online]. Available: https://www.mdpi.com/2079- 9292/11/7/988

[9] A. Almjally and W. S. Almukadi, "Deep computer vision with artificial intelligence based sign language recognition to assist hearing and speech-impaired individuals," *Scientific Reports*, vol. 15, no. 1, p. 32268, 2025. [Online]. Available: https://doi.org/10.1038/s41598-025-09106-8

[10] J. Shin, M. A. M. Hasan, A. S. M. Miah, K. Suzuki, and K. Hirooka, "Japanese sign language recognition by combining joint skeleton-based handcrafted and pixel-based deep learning features with machine learning classification," *CMES - Computer Modeling in Engineering and Sciences*, vol. 139, no. 3, pp. 2605–2625, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1526149224001127

[11] A. KASAPBAS¸ I, A. E. A. ELBUSHRA, O. AL-HARDANEE, and A. YILMAZ, "Deepaslr: A cnn based human computer interface for american sign language recognition for hearing- impaired individuals," *Computer Methods and Programs in Biomedicine Update*, vol. 2, p. 100048, 2022. [Online]. Available:

https://www.sciencedirect.com/science/article/pii/S2666990021000471

[12] D. K. Singh, "3d-cnn based dynamic gesture recognition for indian sign language modeling," *Procedia Computer Science*, vol. 189, pp. 76–83, 2021, aI in Computational Linguistics. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050921011650

[13] L. A. Hussein and Z. S. Mohammed, "Arslr-ml: A python-based machine learning application for arabic sign language recognition," *Software Impacts*, vol. 24, p. 100746, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2665963825000065

[14] N. Salem, S. Alharbi, R. Khezendar, and H. Alshami, "Real-time glove and android application for visual and audible arabic sign language translation," *Procedia Computer Science*, vol. 163, pp. 450–459, 2019, 16th Learning and Technology Conference 2019Artificial Intelligence and Machine Learning: Embedding the Intelligence. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050919321672

[15] K. Inoue, T. Shiraishi, M. Yoshioka, and H. Yanagimoto, "Depth sensor based automatic hand region extraction by using time-series curve and its application to japanese finger- spelled sign language recognition," *Procedia Computer Science*, vol. 60, pp. 371–380, 2015, knowledge-Based and Intelligent Information Engineering Systems 19th Annual Conference, KES- 2015, Singapore, September 2015 Proceedings. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050915022723