Sign Language Recognition using Neural Networks

Arpita Agarwal[1],Dheeraj Devadas[2], Aditya Gupta[3], Ayush Duduskar[4] and Gaurav Patil[5]
[1]*Faculty, Department of Computer Engineering, Pillai College of Engineering, New Panvel*
[2345]*Student Member, Department of Computer Engineering, Pillai College of Engineering, New Panvel*

***Abstract*— Sign Language Recognition (SLR) systems have become increasingly vital in facilitating communication between deaf and hearing communities. While traditional SLR approaches relying on handcrafted features and sensor-based inputs face limitations in scalability and real-time performance. Recent breakthroughs in deep learning have significantly advanced the capabilities of the field. This survey paper comprehensively examines the evolution of SLR systems, from conventional methods to advanced learning models such as CNNs, RNNs, LSTMs, and Transformers. This paper analyzes how these models tackle important problems in both static and dynamic gesture recognition, with particular attention to their computational requirements and performance trade-offs.**

Building on this foundation, we highlight emerging hybrid approaches that combine CNNs for spatial feature extraction with RNNs/LSTMs for temporal modelling - a methodology we are implementing in our ongoing work for continuous sign language recognition. The survey explores key challenges in continuous SLR, including challenges like segmenting gestures, managing blending between signs, and integrating hand, facial, and body cues for multimodal recognition. We further examine recent advancements in real-time processing and accessibility features such as regional language translation. Through this comprehensive review, we identify current limitations in the field and propose future research directions to develop more robust, efficient, and inclusive SLR systems that can operate effectively in diverse real-world conditions.

***Index Terms*— SLR, Artificial Intelligence, Neural Computation, CNNs, RNNs, LSTMs, Hybrid Recognition, Multi-Modal Systems, Recurrent Neural Networks, Long-Short Term Memory Neural Networks, Multimodal Recognition**

I. INTRODUCTION

Sign Language Recognition (SLR) has evolved into a major area within assistive technologies, aiming to facilitate interaction between individuals with hearing impairments and the general population. With over 70 million deaf individuals globally relying on sign language as their primary mode of communication, the need for accurate, real-time, and scalable SLR systems has never been more pressing. Traditional approaches, such as sensor-based gloves or handcrafted feature extraction, have been limited by poor generalization, high costs, and lack of signer independence. With the emergence of deep learning, the landscape of SLR has undergone a significant transformation SLR, enabling systems to interpret both isolated signs and continuous sentences with unprecedented accuracy.

This survey paper provides a comprehensive analysis of SLR methodologies, focusing on the shift from classical machine learning to modern neural network-based approaches, including CNNs, RNNs, LSTMs, and Transformers. We examine their strengths, limitations, and applicability in real-world scenarios, with particular emphasis on hybrid architectures that combine spatial and temporal modelling—an approach central to our ongoing implementation of a real-time, multimodal SLR system.

A. Objectives

The primary objectives of this survey are:

- Explore how sign language recognition has progressed from early sensor-based systems to modern AI-driven approaches.
- Compare traditional methods with newer deep learning techniques to highlight advancements and limitations.
- Identify common obstacles in creating accurate and real-time sign language recognition systems.
- Discuss issues like signer variability, background noise, and the complexity of continuous signing.
- Explain how combining different neural network architectures (like CNNs and RNNs) improves recognition of both static signs and flowing sentences.
- Connect these insights to our planned project, which will use a similar hybrid approach.
- Showcase how SLR technology can be used in education, healthcare, and daily communication.
- Encourage further research into making this technology accessible to diverse communities worldwide.

II. LITERATURE REVIEW

In recent years, notable progress has been achieved in SLR through the application of both machine learning and deep learning strategies. Replacing manual feature

extraction with end-to-end neural architectures has greatly improved both robustness and scalability of SLR systems. This section reviews recent research papers relevant to the field, categorized by their methodological contributions.

[1] Sign Language Recognition: A Deep Survey (2021)

R. Rastgoo, K. Kiani, S. Escalera

This work deeply analyzes current SLR methods by classifying them based on input modalities (e.g., RGB images, depth maps, skeletal data), model families like CNNs, RNNs, and Transformers, and challenges such as independent signer adaptation, gloss-level segmentation, and model scalability. It serves as a comprehensive reference for researchers exploring neural network approaches in sign language recognition.

[2] SIGNFORMER: DeepVision Transformer for Sign Language Recognition (2023)

D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, S. A. Bahaj

This study introduced SIGNFORMER utilizes a Transformer-based design where input sign images are broken into patches and processed using attention mechanisms for ISL recognition. The system achieved 99.29% accuracy on a custom dataset with fewer training epochs than CNNs. However, its reliance on static signs and computational complexity limits real-time or dynamic gesture deployment.

[3] Sign Language Recognition System using Convolutional Neural Network and Computer Vision (2022)

R. S. L. Murali, L. D. Ramayya, V. A. Santosh

A CNN architecture was trained on 2,000 American Sign Language (ASL) gesture images with preprocessing using HSV filtering and morphological operations. The system attained over 90% accuracy for static gesture recognition but was constrained to limited vocabularies and lacked generalizability to dynamic or real-time use cases.

[4] American Sign Language Recognition for Alphabets using MediaPipe and LSTM (2022)

B. Sundar, T. Bagyammal

This work combined Google MediaPipe for hand landmark extraction with LSTM for sequential modelling of ASL alphabet signs. The system processed over 93,000 hand key point frames and achieved 99% accuracy. While effective for static signs, it was not extended to full-word or continuous SLR scenarios.

[5] An Efficient Sign Language Recognition (SLR) System using Camshift Tracker and Hidden Markov Model (HMM) (2021)

P. P. Roy, P. Kumar, B.-G. Kim

A hybrid approach was presented using Camshift tracking for hand motion and HMM for temporal classification. The model could handle both one- and two-handed gestures with low computational requirements. Nonetheless, However, due to the use of manually designed features, performance dropped in complex visual environments and scalability remained limited.

[6] ML Based Sign Language Recognition System (2021)

K. Amrutha, P. Prabu

This study proposed a traditional machine learning system using convex hull features and KNN for static numeric sign classification. While the approach was lightweight and simple, its accuracy was limited (~65%) and unsuitable for dynamic or real-time applications.

[7] American Sign Language Recognition and Training Method with Recurrent Neural Network (2021)

C. K. M. Lee, K. K. H. Ng, C.-H. Chen, H. C. W. Lau, S. Y. Chung, T. Tsoi

A Leap Motion Controller was used to capture ASL gestures, which were classified using an RNN-based model. The system achieved 99.44% accuracy and was implemented as an educational tool with gamified feedback. However, the model was confined to static single-handed alphabet signs.

[8] Skeleton Aware Multi-Modal Sign Language Recognition (2021)

S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, Y. Fu

The paper introduced SAM-SLR, a model integrating RGB, depth, and 3D skeletal data using CNNs and GCNs. With an accuracy of 98.53% on AUTSL, it demonstrated high performance through spatiotemporal modelling. However, the system's dependence on multi-sensor data limits practical deployment.

[9] A Comprehensive Study on Deep Learning-Based Methods for Sign Language Recognition (2021)

A. Adaloglou, T. Chatzis, I. Papastratis, A. Stergioulas, G. T. Papadopoulos, V. Zacharopoulou, G. J. Xydopoulos, K. Atzakas, D. Papazachariou, P. Daras

This paper conducted extensive benchmarking of 2D/3D CNNs, RNNs, and GCNs across multiple datasets, proposing Stimulated CTC and Entropy CTC for gloss prediction. It introduced a novel Greek Sign Language dataset with RGB+D input.

[10] Sign Language Recognition Systems: A Decade Systematic Literature Review (2021)

A. Wadhawan, P. Kumar

This systematic review of 117 SLR papers between 2007–2017 provided insights into modality types, datasets, and algorithmic trends. It highlighted the transition from handcrafted to deep learning models and pointed out research gaps, particularly in real-time and multilingual SLR.

[11] Egyptian Sign Language Recognition using CNN and LSTM (2021)

A. Elhagry, R. G. Elrayes

The authors built a hybrid CNN-LSTM system to classify Egyptian Sign Language gestures. The CNN achieved 90% accuracy, while the hybrid model dropped to 72%, indicating difficulty in optimizing LSTM layers with limited training data and sequence variability.

[12] Fully Convolutional Networks for Continuous Sign Language Recognition (2020)

K. L. Cheng, Z. Yang, Q. Chen, Y.-W. Tai

This study proposed a fully convolutional model for continuous SLR using a Gloss Feature Enhancement (GFE) module. It reached a WER of 3.0% on the CSL dataset without using RNNs. Despite its strengths, precise alignment between glosses and signs remained a challenge.

[13] Sign Language Transformers: Joint End-to-End Sign Language Recognition and Translation (2020)

N. C. Camgoz, O. Koller, S. Hadfield, R. Bowden

The paper introduced a Transformer-based model for both sign recognition and translation, trained on gloss-free sequences using CTC loss. It achieved state-of-the-art BLEU scores on RWTH-PHOENIX-Weather 2014T but required large, annotated datasets and significant computational resources.

[14] Word-Level Deep Sign Language Recognition from Video: A New Large-Scale Dataset and Methods Comparison (2020)

D. Li, C. Rodriguez, X. Yu, H. Li

Li et al. proposed Pose-TGCN and released the WLASL dataset with over 21,000 ASL word-level videos. This graph-based method leveraged hand keypoints to learn temporal dynamics effectively. However, relying solely on pose information proved inadequate when handling variations in signers and lengthy gesture sequences.

[15] Deep Learning-Based Sign Language Recognition System for Static Signs (2020)

A. Wadhawan, P. Kumar

The authors designed a CNN model trained on grayscale representations of Indian Sign Language to detect isolated static signs. With an accuracy nearing 99.9%, the model was effective for discrete gestures but lacked flexibility for sequence-based or continuous sign interpretation.

III. EXISTING SYTEM

Existing Sign Language Recognition (SLR) systems have evolved from traditional handcrafted techniques to deep learning-based approaches. The primary objective of these systems is to bridge the communication gap between the hearing-impaired community and the public by converting visual sign inputs into spoken or written language forms.

Early systems relied heavily on rule-based models, skin colour segmentation, and motion tracking methods. For instance, some methods employed Hidden Markov Models (HMMs) to detect and classify hand trajectories based on predefined gesture patterns. Although lightweight and real-time, these systems often struggled with complex backgrounds, occlusions, and variations in lighting conditions, making them less reliable in uncontrolled environments.

Deep learning advancements positioned CNNs as the backbone of image-based gesture classifiers, particularly effective for recognizing static signs. These CNN models learn spatial characteristics from hand visuals and use dense output layers to assign gesture classes.

RNNs and LSTM networks were adopted to manage time-sequenced characteristics in dynamic sign processing. These models are capable of modelling sequential dependencies across frames in a sign gesture. For example, systems that combine LSTM with pose estimation frameworks like MediaPipe are effective for recognizing alphabets, word and phrases by extracting hand landmarks and passing them into temporal models for classification.

Recent advances explore Transformer-based architectures, which leverage attention mechanisms for learning long-range dependencies in continuous sign language videos. These models achieve state-of-the-art results on benchmark datasets but typically require large, annotated corpora and significant computational resources.

Despite these advancements, existing systems still face several challenges:

- Lack of robustness to signer variability (skin colour, hand size, signing style)
- Limited datasets for less commonly used sign languages
- High computational demand for real-time deployment on mobile or embedded devices
- Absence of multi-lingual or context-aware SLR capabilities

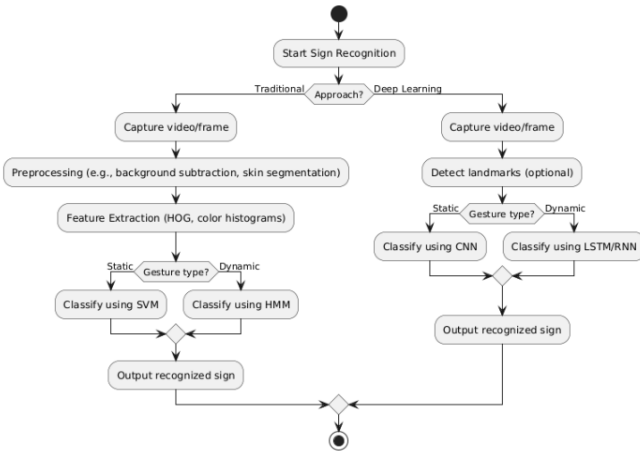

Figure 1. Existing System Activity Diagram

IV. PROPOSED SYSTEM

Our proposed system represents a significant advancement in Sign Language Recognition (SLR) technology by developing a comprehensive, real-time multimodal framework that facilitates seamless communication between deaf/hard-of-hearing individuals and non-signers. This innovative solution moves beyond conventional SLR systems by implementing a sophisticated hybrid deep learning architecture that combines the strengths of multiple neural network paradigms. The system uniquely integrates three critical modalities - precise hand gesture tracking, facial expression analysis, and upper body posture recognition - to achieve unprecedented accuracy in sign language interpretation. At its core, the architecture employs Convolutional Neural Networks (CNNs) for robust spatial feature extraction from individual frames, coupled with Long Short-Term Memory (LSTM) networks for modeling the temporal dynamics of continuous signing. A distinctive feature of our system is its advanced translation module that not only converts recognized signs into grammatically correct sentences but also provides support for multiple regional languages, including Hindi and Tamil, significantly enhancing its practical utility in diverse linguistic contexts. The system's modular design philosophy ensures easy extensibility to incorporate additional sign languages, more complex gestures, or integration with mobile platforms, making it a future-proof solution for inclusive communication technology.

A. Implementation Workflow

The proposed system operates through a structured, multi-stage pipeline designed for efficient and context-aware gesture recognition and translation:

1. **Real-Time Input Acquisition**

The system begins by capturing a continuous video stream from a webcam. This video feed forms the foundation for gesture recognition and is processed frame-by-frame to ensure low-latency, real-time responsiveness. It supports both RGB input to enhance gesture detection under varying environmental conditions.

2. **Hand and Body Landmark Detection**

Using the MediaPipe framework, the system extracts key landmarks from hands, face, and upper body. These landmarks represent the spatial coordinates of joints and facial features, which are then converted into numerical vectors that serve as input for gesture classification. This approach ensures robustness to lighting, background variations, and signer appearance.

3. **Static vs. Dynamic Gesture Classification**

The system distinguishes between static and dynamic gestures:

- Static Gestures are recognized from a single frame using a CNN trained on labelled static gestures. The output is a high-confidence prediction over predefined classes.
- Dynamic Gestures, such as words or phrases, require analysing multiple consecutive frames. These sequences are processed using LSTM models, which learn temporal patterns and signing styles to recognize gesture sequences accurately.

4. **Multimodal Context Integration**

To enhance accuracy and semantic understanding, the system fuses information from hand movements, facial expressions, and body posture. An attention mechanism weighs the importance of each modality dynamically, improving recognition especially in signs that have similar hand shapes but differ in emotional tone or context.

5. **Sentence Formation and Grammar Refinement**

Recognized signs are assembled into coherent sentences using a language model. This stage performs contextual smoothing, syntactic correction, and optionally translates the sentence into a regional language. This ensures the output is both linguistically accurate and user-friendly.

6. **Real-Time Output Generation**

The final translated output is delivered in two formats:

- **Text Display**: Rendered live on-screen for visual confirmation.
- **Speech Synthesis**: A Text-to-Speech (TTS) engine vocalizes the sentence, enabling communication with non-signers.

This modular architecture enables the system to deliver high recognition accuracy, natural sentence construction, and regional adaptability in real time—positioning it as a comprehensive and inclusive solution for sign language interpretation.

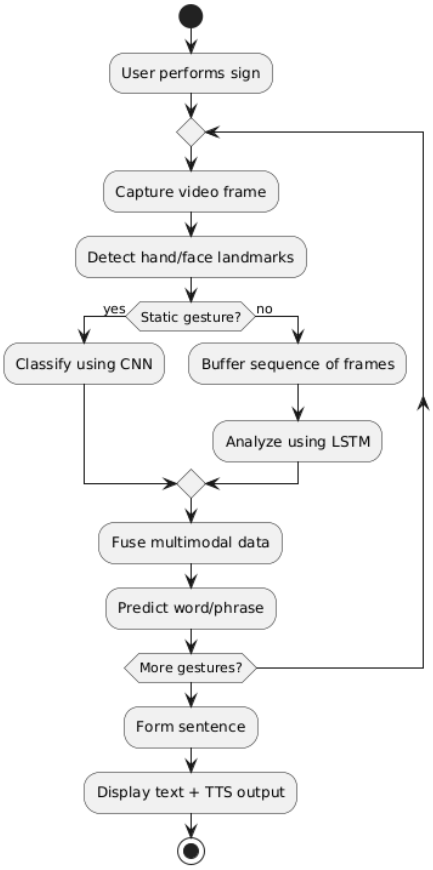

Fig 2. Proposed System Acitivity Diagram

V. CONCLUSION

Sign language recognition has emerged as a vital assistive technology, transforming how deaf and hearing communities interact. This survey has examined the remarkable evolution of SLR systems from early sensor-based methods to today's advanced deep learning approaches. While traditional techniques struggled with scalability and real-time performance, modern neural networks - particularly hybrid CNN-RNN architectures and transformer models - have demonstrated unprecedented accuracy in recognizing both isolated signs and continuous sentences. Our analysis highlights how integrating multimodal inputs (hand gestures, facial expressions, and body posture) significantly improves recognition capabilities, as evidenced by the system we are implementing. However, challenges persist in computational efficiency, dataset diversity, and real-world deployment. The field now stands at a crucial juncture where technical innovations must be balanced with practical usability considerations to create truly inclusive communication solutions. This survey serves not only to document these advancements but also to emphasize the growing importance of developing SLR systems that are accurate, accessible, and adaptable to diverse user needs across different languages and cultures.

VI. FUTURE SCOPE

Looking ahead, several promising directions could further advance sign language recognition technology. First, developing lightweight yet powerful models optimized for mobile and edge devices would enable wider real-world adoption, particularly in resource-constrained settings. There's a significant need for more comprehensive and diverse datasets that capture various signing styles, regional variations, and demographic factors to improve system generalization. Future work should focus on enhancing continuous sign recognition capabilities through better temporal modeling and context-aware algorithms that understand complete sentences rather than isolated signs. The integration of additional modalities like eye tracking and lip reading could provide richer contextual understanding. Another critical area is the development of more sophisticated translation systems that not only convert signs to text/speech but also adapt outputs based on user preferences and regional language nuances. Furthermore, creating bidirectional communication systems that can also generate sign language animations from speech/text would make interactions more natural. As technology matures, establishing standardized evaluation metrics and conducting extensive user studies with the deaf community will be essential to ensure these systems meet real-world needs. Finally, exploring applications in augmented and virtual reality could open new possibilities for immersive learning and communication experiences. These advancements, combined with ongoing improvements in AI hardware and algorithms, position SLR technology to make transformative impacts on accessibility and inclusion in the coming years.

VII. ACKNOWLEDGEMENT

We sincerely thank our faculty members, mentors and our guide Prof. Arpita Agarwal who generously shared their time and expertise, providing invaluable guidance that helped shape this research. Their insightful feedback and encouragement were instrumental in refining this survey paper.

We gratefully acknowledge the authors whose contributions in this domain laid the groundwork for our review. Their contributions have been crucial in advancing sign language recognition systems, and we have endeavoured to represent their findings accurately and respectfully.

Lastly, we acknowledge the deaf and hard-of-hearing community, whose lived experiences and communication needs continue to inspire and motivate our work in developing inclusive technologies. We hope this survey contributes meaningfully to ongoing efforts in bridging communication gaps through technological innovation.

REFERENCES

[1] R. Rastgoo, K. Kiani, and S. Escalera, "Sign language recognition: A deep survey," *Expert Systems with Applications*, vol. 164, p. 113794, 2021.

[2] D. R. Kothadiya, C. M. Bhatt, T. Saba, A. Rehman, and S. A. Bahaj, "SIGNFORMER: deepvision transformer for sign language recognition," *IEEE Access*, vol. 11, pp. 4730–4739, 2023.

[3] R. S. L. Murali, L. D. Ramayya, and V. A. Santosh, "Sign language recognition system using convolutional neural network and computer vision," *Int. J. Eng. Innov. Adv. Technol.*, vol. 4, pp. 138–141, 2022.

[4] B. Sundar and T. Bagyammal, "American sign language recognition for alphabets using MediaPipe and LSTM," *Procedia Computer Science*, vol. 215, pp. 642–651, 2022.

[5] P. P. Roy, P. Kumar, and B.-G. Kim, "An efficient sign language recognition (SLR) system using Camshift tracker and hidden Markov model (HMM)," *SN Computer Science*, vol. 2, no. 2, p. 79, 2021.

[6] K. Amrutha and P. Prabu, "ML based sign language recognition system," in *Proc. Int. Conf. Innovative Trends in Information Technology (ICITIIT)*, 2021, pp. 1–6.

[7] C. K. M. Lee, K. K. H. Ng, C.-H. Chen, H. C. W. Lau, S. Y. Chung, and T. Tsoi, "American sign language recognition and training method with recurrent neural network," *Expert Systems with Applications*, vol. 167, p. 114403, 2021.

[8] S. Jiang, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, "Skeleton aware multi-modal sign language recognition," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 3413–3423.

[9] A. Adaloglou et al., "A comprehensive study on deep learning-based methods for sign language recognition," *IEEE Transactions on Multimedia*, vol. 24, pp. 1750–1762, 2021.

[10] A. Wadhawan and P. Kumar, "Sign language recognition systems: A decade systematic literature review," *Archives of Computational Methods in Engineering*, vol. 28, pp. 785–813, 2021.

[11] A. Elhagry and R. G. Elrayes, "Egyptian sign language recognition using CNN and LSTM," *arXiv preprint arXiv:2107.13647*, 2021.

[12] K. L. Cheng, Z. Yang, Q. Chen, and Y.-W. Tai, "Fully convolutional networks for continuous sign language recognition," in *Computer Vision – ECCV 2020*, pp. 697–714.

[13] N. C. Camgoz, O. Koller, S. Hadfield, and R. Bowden, "Sign language transformers: Joint end-to-end sign language recognition and translation," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10023–10033

[14] D. Li, C. Rodriguez, X. Yu, and H. Li, "Word-level deep sign language recognition from video: A new large-scale dataset and methods comparison," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, 2020, pp. 1459–1469.

[15] A. Wadhawan and P. Kumar, "Deep learning-based sign language recognition system for static signs," *Neural Computing and Applications*, vol. 32, no. 12, pp. 7957–7968, 2020.