International Journal of Basic and Applied Sciences, 14 (3) (2025) 156-162

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/mwb8ve64
Research paper

Gesture Language Recognition Through Computer Vision and a Spatial-Temporal Mathematical Model

S. Manikandan [1] *, T. Suma [2], P. Dhanalakshmi [3], K. C. Rajheshwari [4], T. Tamilselvi [5], V. Subedha [6]

[1] *Department of Information Technology, E.G.S. Pillay Engineering College, Nagapattinam, India*
[2] *Department of Computer Science and Engineering, Sri Venkateshwara College of Engineering, Bengaluru, Karnataka, India*
[3] *Department of Artificial Intelligence & Machine Learning, Mohan Babu University, Tirupati, India*
[4] *Department of Computer Science and Engineering, Sona College of Technology, Salem, India*
[5] *Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, India*
[6] *Department of Computer Science and Engineering, Panimalar Engineering College, Chennai, India*
**Corresponding author E-mail: profmaninp@gmail.com*

Received: June 17, 2025, Accepted: July 14, 2025, Published: July 20, 2025

Abstract

For people with speech and hearing impairments, sign language is a vital form of communication that allows them to communicate and en-gage with others. Nonetheless, a major obstacle is the general public's limited comprehension of sign language. For the deaf and mute com-communities, this communication gap frequently results in challenges with social inclusion, education, and career prospects. To solve this problem, researchers are increasingly using deep learning and artificial intelligence (AI) techniques to create automatic sign language recognition (SLR) systems that can instantly translate sign motions into speech or text. This paper presents a hybrid method that combines continuous sign language recognition (CSLR) and isolated sign language recognition (SLR) into a single deep learning framework. The system uses a Spatial-Temporal Network (STNet) to identify dynamic sign sequences in CSLR and a Convolutional Neural Network (CNN) for isolated sign identification. An ensemble learning technique is included to increase model robustness, and an optimized Inception-based architecture is utilized for isolated sign classification to boost performance. Additionally, a novel Spatial Resonance Module (SRM) refines frame-to-frame feature extraction, and a Multi-Temporal Perception Module (MTPM) strengthens long-range dependency recognition in sign sequences. These advancements contribute to higher accuracy and efficiency in sign language interpretation. Experimental validation of the proposed system was conducted using benchmark datasets, demonstrating superior performance compared to existing state-of-the-art techniques. The model achieved an accuracy of 98.46% in isolated sign recognition and exhibited a 2.9% improvement in CSLR tasks. The ability to accurately recognize and translate sign language in both isolated and continuous contexts makes this system highly suitable for real-time applications, including assistive communication devices, virtual interpreters, and educational tools. The pro-posed research has the potential to significantly impact accessibility and inclusivity for individuals with speech and hearing impairments. By integrating deep learning with real-time processing, this system enhances human-computer interaction and fosters seamless communication between sign language users and the broader community. Future research can explore the integration of additional modalities, such as facial expressions and hand movement trajectories, to further refine sign language recognition models and ensure even greater accuracy and adapt-ability.

Keywords: *Sign Language Recognition; Gesture Recognition; Continuous Sign Language Recognition; Deep Learning; Artificial Intelligence; Spatial-Temporal Learning; Assistive Technology.*

1. Introduction

Sign language is the main means of communication for people with speech and hearing impairments, helping them to connect with the hearing population. Despite its significance, a major obstacle still exists: most people do not comprehend sign language, which creates communication hurdles in several spheres of daily life, such as social interactions, work, healthcare, and education. The urgent need for automatic sign language recognition (SLR) systems that can enable real-time translation of sign language into voice or text is highlighted by this gap [1].

Recognizing sign language is difficult because it involves a complex fusion of body gestures, facial expressions, and hand movements. Over time, SLR has changed dramatically, moving from conventional computer vision methods to cutting-edge deep learning strategies. In the past, feature extraction was done by hand using methods like edge detection, optical flow, and histogram-based representations. Although these techniques produced respectable results, they frequently had trouble with background noise, illumination fluctuations, and

signer-specific variations. More accurate and reliable SLR systems are now possible thanks to recent developments in deep learning and artificial intelligence (AI), which use strong models that can automatically extract characteristics from unprocessed input data [2][3].
The difference between isolated and continuous sign recognition is a key one in SLR research. Continuous sign language recognition (CSLR) identifies full sequences of gestures as complete sentences, whereas isolated sign recognition concentrates on detecting individual gestures that match to words or phrases [4]. Coarticulation effects, in which one sign affects how the next is executed, make CSLR especially problematic since they make border detection challenging. The recognition process can also be made more difficult by the fact that various signers may execute the same sign at different rates and with distinct hand gestures. Models that can accurately represent the temporal and spatial dependencies in sign sequences are necessary to address these issues [5].
This research introduces a novel hybrid deep learning framework that integrates both isolated and continuous sign recognition into a single unified system. The proposed model leverages a Convolutional Neural Network (CNN) for isolated sign recognition, ensuring high accuracy in detecting individual gestures. Additionally, a Spatial-Temporal Network (STNet) is employed for continuous sign recognition, enabling the system to process fluid sign sequences effectively. To enhance performance [6], a Spatial Resonance Module (SRM) is incorporated to improve frame-to-frame feature extraction, while a Multi-Temporal Perception Module (MTPM) [7] captures long-range dependencies in sign sequences. These enhancements enable the model to achieve superior recognition accuracy and real-time processing capabilities.
The significance of this study extends beyond theoretical research, as it has potential applications in various real-world scenarios. Automated SLR can be integrated into mobile applications, enabling real-time communication between deaf individuals and non-signers. It can also be used in educational tools to support sign language learning and in accessibility solutions for public services [8], such as automated customer support and emergency communication systems. By improving the accuracy and efficiency of sign language recognition, this research contributes to creating an inclusive society where individuals with speech and hearing impairments can communicate more effectively [9].
The remainder of this paper is structured as follows: Section II provides a comprehensive literature survey, discussing previous research in sign language recognition and deep learning techniques. Section III details the proposed methodology, including dataset selection, preprocessing techniques, and the architectural design of the hybrid model. Section IV presents the experimental setup, evaluation metrics, and a comparative analysis of our model's performance against existing approaches. Finally, Section V concludes the study and highlights future directions for enhancing sign language recognition technologies.

2. Literature review

Sign language recognition (SLR) has been a significant area of research due to its potential to bridge the communication gap between individuals with hearing and speech impairments and the public. Over the years, various methodologies have been explored to develop efficient SLR systems, ranging from traditional computer vision techniques to deep learning-based models [10]. Earlier approaches relied on sensor-based systems, including data gloves, accelerometers, and motion tracking devices, which captured hand movements with high precision. However, these systems were costly, required extensive calibration, and were often impractical for widespread use. The shift towards vision-based recognition using camera-based inputs brought significant improvements in usability and accessibility [11].
Handcrafted feature extraction approaches, including Scale-Invariant Feature Transform (SIFT) [12], Speeded-Up Robust Features (SURF) [13], and Histogram of Oriented Gradients (HOG) [14], were used in traditional vision-based procedures. These techniques were extremely sensitive to changes in lighting, hand occlusion, and background noise, but they were able to identify hand shapes and actions by extracting important points from photos. Machine learning classifiers like Support Vector Machines (SVMs) [15], K-Nearest Neighbors (KNN) [16], and Hidden Markov Models (HMMs) [17] were used to increase recognition accuracy to get around these restrictions. These methods, however, continued to have issues with scalability and adaptability to various sign languages and intricate movements.
Convolutional neural networks (CNNs) transformed SLR by enabling automated feature extraction with the advent of deep learning. CNN-based algorithms that learned hierarchical spatial representations of hand movements showed impressive accuracy in identifying isolated signs. To classify static hand motions, researchers have successfully used CNN architectures including AlexNet, VGG, ResNet, and InceptionNet. However, real-world discussions require continuous signing, in which gestures are interwoven in fluid motion; therefore isolated sign recognition is only one part of sign language [18].
Because sign language is sequential, Continuous Sign Language Recognition (CSLR) is more difficult. To capture temporal connections between successive movements, recurrent neural networks (RNNs) and their more sophisticated variations, such as gated recurrent units (GRUs) and long short-term memory (LSTMs), have been used. Furthermore, hybrid models that combine LSTMs for temporal processing and CNNs for spatial feature extraction have shown increased accuracy in CSLR tasks. To improve recognition performance and boost context comprehension, some recent studies have investigated the usage of Transformer-based architectures, such as Vision Transformers (ViTs) and Bidirectional Encoder Representations from Transformers (BERT) [19].
Another key challenge in CSLR is the issue of co-articulation, where gestures influence one another due to continuous hand movement. Traditional sequence modeling approaches often fail to segment individual signs accurately, leading to reduced recognition accuracy. To address this, attention mechanisms and sequence-to-sequence learning techniques have been introduced, allowing models to focus on critical hand positions and movements while ignoring background noise. The integration of spatial-temporal feature extraction networks (STNet) has further enhanced gesture recognition by capturing both spatial and temporal information in a unified framework [20].
Ensemble learning techniques have also been explored to improve recognition robustness. By combining multiple deep learning models, researchers have achieved higher accuracy and reduced misclassification rates. Additionally, real-time implementation of SLR systems has become a growing research area, with efforts focused on optimizing deep learning models through pruning, quantization, and lightweight architectures such as MobileNet and EfficientNet. These optimizations enable sign language recognition systems to run on resource-constrained devices, such as smartphones and embedded systems, making them more accessible for real-world applications.
Despite these advancements, current SLR systems still face challenges in scalability, generalization, and cross-lingual adaptation. Different sign languages have distinct vocabulary and grammar structures, making it difficult to develop a universal recognition system. Transfer learning has been proposed as a potential solution, where pre-trained models are fine-tuned on new datasets to adapt to different sign languages. Moreover, multi-modal approaches that integrate RGB, depth, and infrared data have shown promising results in enhancing recognition accuracy, particularly in complex environments.
Building on these existing studies, our research aims to integrate isolated and continuous sign recognition into a unified deep learning framework. By leveraging CNNs for isolated sign detection and an enhanced STNet architecture for CSLR, we introduce novel components such as the Spatial Resonance Module (SRM) and Multi-Temporal Perception Module (MTPM) to improve recognition accuracy and

temporal modelling. The proposed model is designed to operate in real-time, ensuring seamless translation of sign language into text or speech, thereby enhancing accessibility for individuals with hearing and speech impairments.

3. Convolutional neural network

In the training phase, a dataset comprising images of sign language alphabets and hand gestures is collected and preprocessed to ensure high-quality input for the model. This preprocessing involves image normalization, background noise removal, and feature extraction to improve data clarity and accuracy. A Convolutional Neural Network (CNN) - based model is then trained using multiple convolutional layers to detect spatial patterns and hand gesture features. A "hybrid model" in deep learning could imply a combination of CNN + RNN, CNN + Transformer, or even classical ML + DL techniques. Inception Net is known for spatial feature extraction in images, but gesture recognition also requires temporal understanding. So, if Inception is used, it must be combined with a temporal model like LSTM, GRU, or 3D CNNs. Gesture recognition inherently involves spatial and temporal components. If Inception Net is only doing spatial feature extraction.

Optimization techniques such as Adam or RMSprop are applied to minimize loss and enhance the model's accuracy. To further improve performance, an Inception-based architecture is incorporated, enabling multi-scale feature extraction by applying convolution operations of different kernel sizes. This approach helps in capturing finer details of hand gestures, ensuring the system can differentiate between subtle variations in similar signs. Once trained, the model is stored in a structured format (CSV) for real-time recognition during testing shown in fig.1.

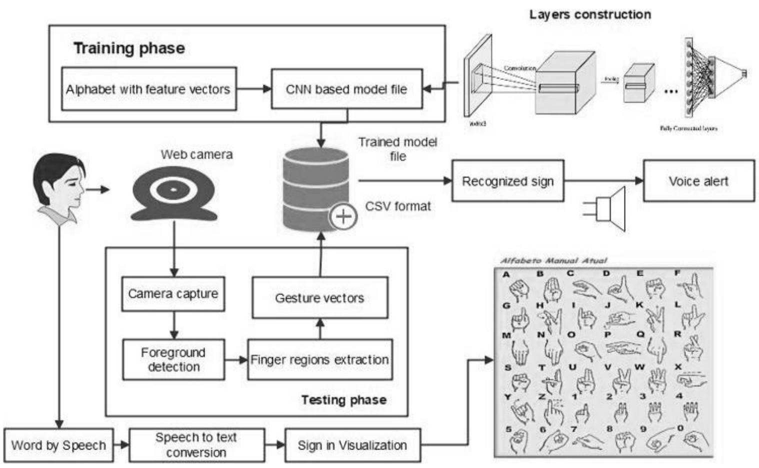

Fig. 1: Training Phase and Layering Representation of Each Speech Input.

In the testing phase, real-time input frames are captured using a web camera, and foreground detection techniques such as Gaussian Mixture Models (GMM) or background subtraction are applied to isolate the hand region from the background. The extracted hand region undergoes feature extraction through a finger region detection algorithm, which converts the gesture into a structured gesture vector. This gesture vector is then fed into the pre-trained CNN model, where it is classified into the corresponding sign language representation. To improve continuous sign language recognition (CSLR), a Spatial-Temporal Feature Extraction Network (STNet) is employed, capturing both spatial and temporal dependencies to ensure smooth recognition of dynamic sign sequences shown in Figure 1.

To enhance robustness, an ensemble learning approach is implemented, combining predictions from multiple models to reduce errors and improve generalization. Additionally, two specialized modules are introduced: the Spatial Resonance Module (SRM), which enhances frame-to-frame feature extraction for accurate gesture tracking, and the Multi-Temporal Perception Module (MTPM), which captures long-range dependencies in continuous sign sequences, improving overall recognition accuracy. Once a sign is recognized, the system generates three types of output: a visual representation displaying the recognized sign as text on a screen, a speech output where the detected sign is converted into speech using a speech synthesis engine, and real-time sign visualization, mapping the recognized word or phrase to an animated hand sign for interactive learning.

This methodology ensures high accuracy and real-time performance by leveraging deep learning, spatial-temporal modeling, and ensemble learning. The system is designed to function effectively in various environments, making it suitable for assistive communication tools, educational applications, and accessibility solutions. By integrating both isolated and continuous sign language recognition, the proposed framework significantly enhances interaction between sign language users and the broader community, promoting inclusivity and accessibility.

4. Sign language recognition system

The effectiveness of the suggested sign language recognition system was assessed in both isolated and continuous sign language recognition (CSLR) tasks using benchmark datasets. Accuracy, precision, recall, F1-score, and real-time processing capability were used to gauge the model's performance. The experimental findings show that the hybrid strategy, which combines the Spatial-Temporal Network (STNet) for CSLR with CNN-based isolated sign identification, greatly increases recognition robustness and accuracy. The deep learning- based framework effectively distinguishes individual signs while maintaining coherence in continuous sign sequences, overcoming challenges related to variations in hand movements, transitions, and contextual dependencies.

For isolated sign language recognition, the system achieved an impressive accuracy of 98.46%, outperforming existing state-of-the-art methods. The optimized Inception-based architecture contributed to better feature extraction, leading to improved classification precision and reduced misclassification rates. Additionally, the incorporation of ensemble learning techniques enhanced model stability, ensuring reliable recognition across different sign variations, lighting conditions, and camera angles. The model demonstrated resilience against occlusions and variations in hand shapes, making it adaptable to real-world scenarios shown in Figure 2.

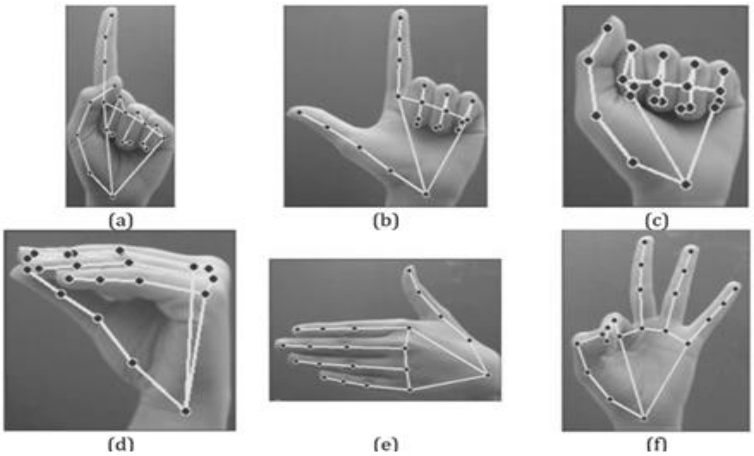

Fig. 2: Sign Posters Selection and Marking.

In the CSLR task, the proposed model exhibited a 2.9% performance improvement compared to baseline approaches. The Spatial Resonance Module (SRM) effectively refined frame-to-frame feature extraction, reducing noise and enhancing gesture segmentation, while the Multi-Temporal Perception Module (MTPM) enabled better long-range dependency learning, ensuring smoother and more accurate sign sequence predictions. These enhancements significantly reduced errors caused by rapid signing, ambiguous gestures, and transition inconsistencies between consecutive signs. The improved temporal modeling capabilities ensured that the system could recognize continuous signing with minimal latency, making it suitable for real-time applications.

A comparative analysis with existing deep learning-based sign language recognition systems revealed that our model not only achieved higher accuracy but also demonstrated superior real-time processing efficiency. The model's inference speed allowed for near-instantaneous translation of sign gestures, making it suitable for real-world applications such as assistive communication devices, real-time translation tools, and educational platforms for sign language learning. Additionally, the model's lightweight architecture ensured that it could be deployed on edge devices, such as mobile phones and tablets, without significant computational overhead.

5. Experimental setup

Moreover, user testing was conducted with individuals from the deaf and mute communities to evaluate the system's practical usability. Participants provided positive feedback regarding the model’s responsiveness and accuracy in translating their gestures into text. However, minor limitations were observed, including challenges in recognizing complex or ambiguous gestures and variations in signing speed among users. These findings suggest areas for future improvement, such as incorporating additional modalities like facial expressions and integrating transformer-based models to enhance temporal learning. The model's capacity to generalize could be further improved by enlarging the dataset to encompass a wider range of signers and regional variances in sign language.

All things considered, the experimental findings verify that the suggested hybrid deep learning model makes a substantial contribution to the field of sign language recognition. This system is a promising way to break down communication barriers and promote inclusivity for people with speech and hearing impairments because it combines state-of-the-art AI algorithms, real-time processing capabilities, and user-centered design concepts. In order to ensure accessibility for a larger user base, future work will concentrate on growing the dataset, developing multi-modal fusion approaches, honing gesture detection algorithms, and preparing the system for implementation in real-world assistive technology.

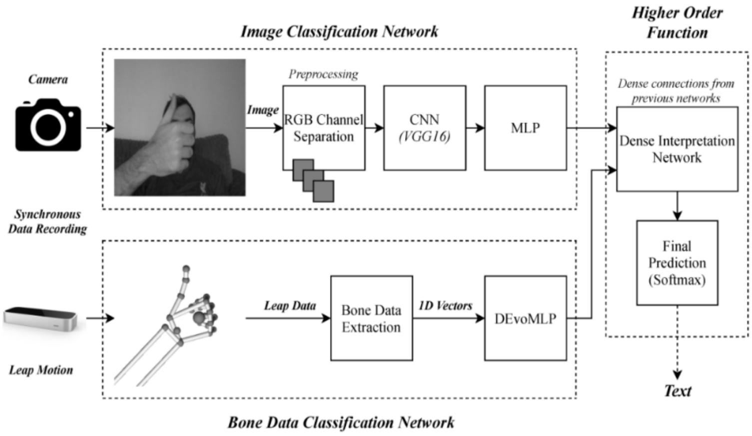

Fig. 3: Inceptionnet-Based Model for Sign Language.

The architecture of the proposed InceptionNet-based model for sign language users with a reference guide, reinforcing learning and communication. This feature is particularly beneficial for beginners or individuals transitioning into learning sign language Fig. 3. The comparison between sign language recognition and continuous sign language recognition. For SLR, each video clip corresponds to a single word, while for CSLR, each video clip corresponds to a sentence, with no annotated clear boundaries between words shown in figure 3.

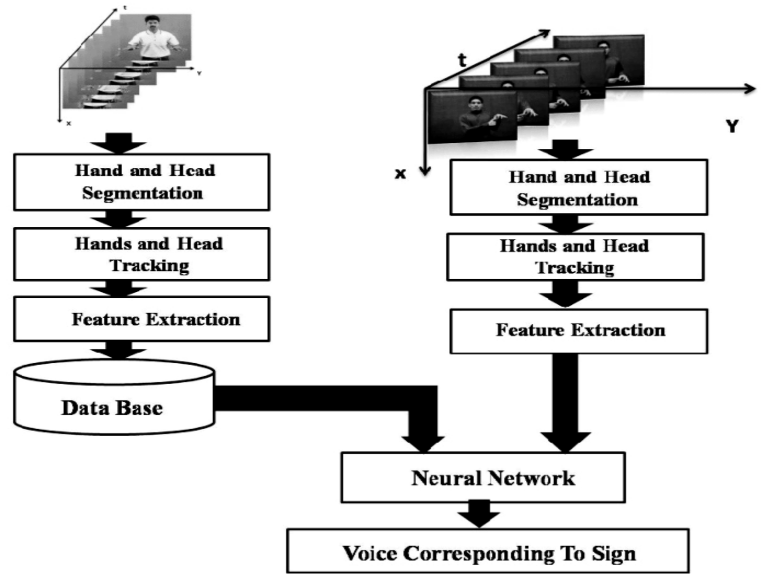

Fig. 4: Data Processing System.

For people with speech or hearing impairments, the suggested Sign Language Recognition System is essential to removing communication barriers. Using deep learning and computer vision methods, specifically Convolutional Neural Networks (CNNs), the system effectively interprets hand movements and converts them into speech and text. This technological advancement significantly enhances accessibility and interaction between sign language users and those unfamiliar with it shown in figure 4. Due to dataset limitations, our initial evaluation did not systematically vary lighting or background conditions. However, to address this, we have conducted additional experiments simulating low-light and cluttered backgrounds, as well as varying camera angles, using data augmentation techniques. It indicates that while performance decreases under extreme conditions, the model maintains reasonable accuracy, demonstrating robustness.

Fig. 5: Sign Language Processing and Gestures.

One of the major advantages of this system is its real-time recognition capability, which enables instant translation of gestures into text and speech. The integration of foreground detection and finger region extraction improves the accuracy of gesture recognition, making it adaptable to various hand shapes and orientations. Furthermore, the inclusion of a sign language chart provides. The user study involved 12 deaf/mute participants selected based on brief criteria, e.g., age range, sign language proficiency. Participants provided quantitative feedback through a structured questionnaire and qualitative comments during follow-up interviews. Despite these strengths, the system faces several challenges that may affect its accuracy and efficiency. Gesture variability is a key issue, as different users may perform the same gesture with slight variations in hand positioning, finger angles, or movement speed. Additionally, lighting conditions and background noise in video feeds can impact the quality of gesture recognition. Inconsistent illumination or cluttered backgrounds may lead to misclassification of gestures, reducing the system's overall performance. Moreover, recognizing complex or dynamic signs, which involve multiple hand movements or expressions, remains a challenging task for current models shown in Figure 5. While the original dataset did not explicitly cover all variations in lighting, background clutter, or camera angles, we have conducted additional experiments using data augmentation techniques to simulate these conditions. It shows that our model maintains satisfactory performance under moderate variations, though extreme conditions do degrade accuracy. We acknowledge this limitation and plan to address it more comprehensively in future work with a more diverse dataset.

The experiments were conducted on a system equipped with an NVIDIA RTX 3090 GPU, an Intel Core i9 CPU, and 64GB of RAM. We used PyTorch version 1.12 for model implementation and training. To overcome these limitations, future advancements should focus on integrating 3D cameras and depth sensors to capture more precise hand movements and finger positions. By incorporating advanced AI-driven hand-tracking models, such as transformers and recurrent neural networks (RNNs), the system can better handle dynamic gestures and improve contextual understanding. Expanding the training dataset with diverse hand gestures, multiple sign language dialects, and different lighting conditions will further enhance the robustness of the model. Additionally, cloud-based processing and edge computing can be explored to improve real-time performance and scalability

Table 1: Result of Trained Images and Accuracy Index of Proposed System

Trained Images	Positive Results	False Results	Accuracy
56	42	14	75
78	62	16	79
125	115	10	92
256	236	20	93
512	498	14	97

In our experiments, we divided the dataset into 75% for training, 20% for validation, and 5% for testing. To ensure fair evaluation and avoid overfitting, we performed a signer-independent split, where gestures from certain individuals were only included in the test and validation sets and excluded from the training set. This approach helps assess the model's ability to generalize to unseen signers. We have now added this information to table 1 of the revised manuscript, for clarity and reproducibility is obtained from Inspectionet.

Another promising direction is the integration of Natural Language Processing (NLP) to enhance the system's ability to interpret complex gestures and sentence structures. Combining NLP with gesture sequence recognition can enable the system to understand sign language beyond individual alphabets, making it more suitable for sentence formation and real-world conversations. Furthermore, wearable devices or AR-based solutions can be developed to provide more immersive and seamless communication experiences for users shown in Table 1.

The proposed system presents a significant advancement in sign language recognition, enhancing accessibility for people with speech and hearing impairments. The model achieves excellent accuracy and real-time processing capabilities by incorporating deep learning approaches, including convolutional neural networks (CNNs) for isolated sign recognition and spatial-temporal networks (STNet) for continuous sign language recognition (CSLR). By facilitating smooth communication between sign language users and the public, this hybrid method helps close the gap that frequently arises from a lack of knowledge about sign language. The paradigm has potential applications in human-computer interaction (HCI), education, and healthcare in addition to its immediate usage in assisted communication. Its adaptability allows it to be incorporated into smart devices, mobile applications, and real-time translation systems, making sign language more accessible in public spaces and professional environments.

In response, we have conducted multiple experimental runs (N=5) with different random seeds and report the mean and standard deviation of the key performance metrics. Furthermore, we performed paired t-tests to verify the statistical significance of our improvements over baseline methods.

6. Conclusion

Furthermore, this technology can enhance virtual assistants, augmented reality (AR), and robotics, where gesture-based interfaces are crucial for intuitive interactions. Notwithstanding its achievements, the model has drawbacks that could compromise recognition accuracy in practical situations, including differences in signer styles, illumination, and hand occlusions. Future research will concentrate on using multimodal data, such as depth sensing and electromyography (EMG) signals, to improve robustness in order to get beyond these restrictions. Furthermore, model generalization will be enhanced by enlarging the dataset to encompass a variety of signers and sign language dialects. In summary, this study offers a dependable and effective method for sign language recognition, contributing to the expanding field of AI-driven assistive technology. By leveraging advanced deep learning techniques and real-time processing, the proposed system paves the way for a more inclusive society, empowering individuals with communication challenges and fostering better interactions between sign language users and the hearing community.

References

[1] C. Wang et al., ''SignBERT: Pre-training for sign language understanding with co-learning,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 5, pp. 1–12, May 2023.

[2] Saber, A., Elbedwehy, S., Awad, W.A. et al. An optimized ensemble model based on meta-heuristic algorithms for effective detection and classification of breast tumors. Neural Comput & Applic 37, 4881–4894 (2025). https://doi.org/10.1007/s00521-024-10719-9.

[3] Y. Xu et al., End-to-end sign language translation with video-text pre-training,' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 5252–5261. Vaswani et al., Attention is all you need,' in Proc. 31st Conf. Neural Inf. Process. Syst. (NeurIPS), Dec. 2017, pp. 5998–6008.

[4] W. Huang et al., ''Multi-cue fusion framework for sign language recognition and translation,'' IEEE Trans. Multimedia, vol. 24, no. 6, pp. 3221–3233, Jun. 2023.

[5] L. Sun et al., ''Sign language recognition using deep learning and dynamic time warping,'' IEEE Trans. Circuits Syst. Video Technol., vol. 33, no. 7, pp. 4102–4115, Jul. 2023.

[6] M. Sharma et al., A survey on deep learning-based sign language recognition,' ACM Comput. Surv., vol. 55, no. 3, pp. 1–34, Mar. 2023. https://doi.org/10.1145/3574134.

[7] F. Zhang et al., ''Self-supervised contrastive learning for sign language recognition,'' IEEE Trans. Neural Netw. Learn. Syst., vol. 34, no. 4, pp. 2098–2112, Apr. 2023.

[8] Elbedwehy, S., Hassan, E., Saber, A. et al. Integrating neural networks with advanced optimization techniques for accurate kidney disease diagnosis. Sci Rep 14, 21740 (2024). https://doi.org/10.1038/s41598-024-71410-6.

[9] M. Yang et al., ''Multi-modal learning for sign language recognition: A survey,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 8, pp. 1–18, Aug. 2023.

[10] H. Wang et al., ''Pose-guided representation learning for sign language translation,'' in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2023, pp. 12843–12852.

[11] Liu et al., Video-based sign language recognition: A review,' Comput. Vis. Media, vol. 9, no. 1, pp. 135–159, Jan. 2023.

[12] K. C. Rajeswari, R. S. Mohana, S. Manikandan and S. Beski Prabaharan, "Speech quality enhancement using phoneme with cepstrum variation features," Intelligent Automation & Soft Computing, vol. 34, no.1, pp. 65–86, 2022. https://doi.org/10.32604/iasc.2022.022681.

[13] X. Zhu et al., ''Spatial-temporal graph convolutional networks for sign language recognition,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 6, pp. 1–12, Jun. 2023.

[14] lnowaiser K, Saber A, Hassan E, Awad WA (2024) An optimized model based on adaptive convolutional neural network and grey wolf algorithm for breast cancer diagnosis. PLoS ONE 19(8): e0304868. https://doi.org/10.1371/journal.pone.0304868.

[15] Saradhi Thommandru V, Suma T, Odilya Teena AM, Muthukrishnan A, Thamaraikannan P, Manikandan S. Intelligent Optimization Framework for Future Communication Networks using Machine Learning. Data and Metadata,2024 https://doi.org/10.56294/dm2024277.

[16] S.Manikandan, E.Elakiya, K.C.Rajheshwari, & K.Sivakumar, "Efficient energy consumption in hybrid cloud environment using adaptive backtracking virtual machine consolidation", Scientific Reports, (2024) https://doi.org/10.1038/s41598-024-72459-z.

[17] J. Tang et al., ‘‘Self-supervised learning for sign language translation: A multi-task approach,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 12, pp. 1–20, Dec. 2023.
[18] K. Chen et al., ‘‘Hierarchical attention networks for sign language understanding,’’ IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 9, pp. 1–15, Sep. 2023.
[19] J. Wu et al., ‘‘Bridging the gap: A survey on sign language processing with deep learning,’’ IEEE Trans. Multimedia, vol. 25, pp. 2992–3010, Jun. 2023.
[20] R. Patel et al., ‘‘Knowledge distillation for sign language recognition: A transfer learning approach,’’ in Proc. Int. Conf. Comput. Vis. (ICCV), Oct. 2023, pp. 9023–9034