A Survey on Deep Learning and NLP-Based Real-Time Sign Language Recognition and Contextual Conversion

Kruthika D[1], Prajwalanish V[2], Pragath R[3], Mohammed Ajmal Hussain[4], Venugeetha Y[5]
[1,2,3,4,5]Department of CSE, Don Bosco Institute of Technology,
Bengaluru, Karnataka, India
kruthikagowda118@gmail.com, prajwalanish888@gmail.com, saipragath007@gmail.com,
ah0044204@gmail.com, venugeeta@gmail.com

Abstract— Signed language is an essential means of communication means for the Deaf and Hard-of-Hearing (DHH) community. Conventional techniques of signed language recognition tend to suffer from gesture variability, real-time response, and sentence formation accuracy. This paper provides a survey of current techniques in computer vision-based hand gesture recognition, real-time sign translation, and speech integration, with emphasis on deep learning, multimodal AI, and NLP-based methods. The research emphasizes improvements made in transformer-based models, CNN-LSTM architectures, and generative AI towards achieving a complete real-time sign language recognition and speech synthesis system.

Keywords— Sign Language Recognition, Deep Learning, NLP, Gesture Detection, Real-Time Translation, AI-based Accessibility.

1. Introduction

Sign recognition has been investigated deeply in NLP and computer vision and has also proven an effective means of closing the communication gap for speech-impaired and hearing-impaired individuals. According to the World Health Organization (WHO), in the year 2018, there are about 466 million individuals with hearing impairment worldwide, most of whom use sign language as their only means of communication. Accessibility, however, still remains a challenge, and most still cannot use sign language, which makes up a very small percentage of the population. There are about 70 million sign language users, but technology-based solutions for real-time translation are lacking. Deep learning revolutionized sign language recognition with the possibility of attaining high-precision gesture detection, sentence formation, and speech generation. Rule-based models fall behind in gesture variety, illumination, and real-time learning, while CNN-based models excel in image and video feature extraction, allowing precise hand gesture recognition. By removing temporal dependencies, LSTM and Transformer models also improve recognition and enable continuous sign language processing. Additionally, skeletal tracking is enhanced by OpenPose and MediaPipe, enabling greater movement detection for more accurate translations.

Understanding a sophisticated deep learning model that can recognize sign language motions from still photos, recorded movies, and live video feeds is the goal of this survey. The model will convert these motions into speech for hearing accessibility as well as grammatically accurate sentences that are exhibited or displayed on the screen. The Google Speech API and NLP-text-to-speech conversion provide natural-sounding speech output, which simplifies and eases interaction.

With differing degrees of precision, signed language identification technology can be used in a variety of fields:

- Education: Automated sign translation tools can enhance learning accessibility, with models reaching 90-94% accuracy in classroom settings.
- Healthcare: AI-powered sign recognition, with 92–95% accuracy rates in controlled settings, can facilitate doctor-patient communication.

- Customer Service: AI-powered sign recognition can facilitate real-time interactions in retail and service industries, with accuracy exceeding 88%.
- Legal and Government Services: Sign language translation systems can improve accessibility in courtrooms and public offices, maintaining accuracy levels of 85-90%.

Deep learning and NLP have certain benefits over traditional methods, including greater flexibility, real-time processing, and improved context understanding. Deep learning models, unlike traditional models, are able to learn from large amounts of data and enhance their accuracy with time. Sentence generation by NLP ensures natural phrasing, and the translations are intuitive and easy to use. Multimodal AI fusion—vision, speech, and gesture tracking integration—is more interactive in real-time, and sign language recognition is more scalable and accessible. By utilizing deep learning, NLP, and AI speech synthesis, this project seeks to develop an end-to-end sign language recognition system to facilitate easy communication for the hearing impaired.

Future tasks involve dataset growth, real-time tuning, and porting to mobile and embedded platforms, offering an easier digital communication environment.

1.1 Motivation

- Communication barriers experienced by deaf or hard-of-hearing people.
- Limited availability of real-time high-accuracy AI-powered sign models.
- Having difficulties with the range of signs, interpretation of gesture meaning, and organization and construction of sentences.
- Incorporating AI models for spontaneous speech and text translation.

1.2 Challenges in Current Sign Language Recognition

- Gesture Variability – Variations in signing velocities, styles, and hand locations result in recognition errors.
- Real-Time Processing Bottlenecks – In most situations, classical models will experience very high latency and high computation overhead, which will hurt their overall performance.
- Sentence Formation Problems – Current models can comprehend single signs but not context-sensitive sentence formation.
- Speech Integration Limitations – Natural wording modifications are necessitated in live text-to-speech synthesis.

This in-depth review carefully classifies modern models of AI-based sign recognition, focusing on their high potential to be used in applications with real-time translation needs.

2. Related Work

Various techniques have been attempted for signed language identification and translation. This survey categorizes the techniques into:

- Gesture recognition algorithms (Hybrid structures, CNN, SVM).
- Real-time processing methods (MediaPipe, OpenPose, Transformer-based techniques).
- NLP sentence generation models (BERT, GPT, Neural Machine Translation).

2.1 Computer Vision and AI-Based Gesture Recognition

A. Sharma et al. created a CNN-based segmentation technique that is crucial to the recognition of signed gestures because it can accurately detect hand movements and decipher visual input in sign language systems. In image processing, feature extraction methods like Histogram of Oriented Gradients and Local

Binary Patterns are frequently employed to identify important motion patterns and guarantee accurate gesture recognition. The variety of signing styles presents a significant problem for real-time gesture detection, as variations in hand placement, movement speed, and articulation can provide inconsistent recognition outcomes. Furthermore, the quality of object detection and contrast levels in sign language models can be impacted by lighting conditions. In order to overcome these obstacles and increase the resilience of AI-driven sign language translation solutions, sophisticated preprocessing techniques and adaptive learning algorithms are needed [1].

LSTM-based models, which are used for sequence video-based sign recognition and enable AI to process and analyze continuous sign gestures throughout time, were studied by R. Patel. These models are very useful for comprehending dynamic sign language sequences because they are adept at preserving the contextual relationship between movements. By collecting fine-grained hand movements and spatial information, CNN models with sensor-based hand tracking improve gesture recognition even further and guarantee accurate sign detection in a variety of illumination and environmental circumstances. Sign recognition systems become extremely accurate and efficient by fusing CNN's feature learning capabilities with LSTM's sequential learning capabilities. Furthermore, seamless communication is made possible by real-time sign recognition with integrated speech synthesis, which instantly translates identified movements into spoken language. Projects aimed at enhancing accessibility and bridging the gap between spoken communication and sign language users in practical applications will find this integration especially pertinent [2].

2.2 Real-Time AI-Based Sign-to-Speech Conversion

M. Tanwar worked on a CNN-based hand gestures recognition platform is employed to identify sign language movements with high accuracy, integrating speech conversion through Natural Language Processing to enhance communication accessibility. The model utilizes deep learning techniques to extract meaningful features from hand gestures, ensuring precise recognition of different sign movements. By incorporating Google Speech API, the system converts recognized text into speech in real-time, offering natural-sounding feedback that improves user interaction. This integration allows for a seamless transition between signed language and communication, addressing the division between signers and non-signers. Additionally, optimization techniques are applied to improve processing speed and real-time adaptation, ensuring that gesture recognition and speech synthesis operate efficiently without delays. These enhancements significantly contribute to the practicality and usability of AI-driven sign language translation systems [3].

T. Williams researched transformer-based architecture, which is crucial for the identification of signed language since it makes it possible to translate signed motions into the appropriate text. By controlling sequential dependencies, these models guarantee that translations of signed language preserve grammatical accuracy. The system employs OpenPose analysis, which analyzes body motions, CNN to extract visual data, and optical flow techniques to capture motion dynamics in order to increase the accuracy of gesture detection. This combination makes it possible to comprehend sign language expressions better, which improves recognition accuracy. Additionally, the system employs a two-way learning methodology, optimizing both the text-to-sign and sign-to-text translation models to facilitate the construction of natural sentences. The methodology improves communication fluency by guaranteeing bidirectional learning, which increases the flexibility and significance of sign language translation in practical situations [4].

2.3 Other Relevant Papers in Sign Language AI Research

Beyond the base papers, additional research covers:

- Leap Motion-based Myanmar Sign Language Recognition using Deep Learning (gesture tracking innovations). [5]
- Improving Sign Language Recognition with Machine Learning and AI (dataset augmentation for diverse sign detection). [6]
- SignChatAI—Generative AI for Deaf-and-Mute Community Using Indian Sign Language (ISL) (chatbot-driven sign interactions). [7]
- Comprehensive Residual Neural Network with Augmented Data for Sign Language Recognition (enhancing gesture dataset quality). [8]
- Cross-Modal Integration of Video and Text for Continuous Signed Communication (temporal dependency handling in gesture recognition). [9]
- Hybrid InceptionNet-Based Enhanced Architecture for Isolated Signed Language Recognition (deep learning optimizations for sign detection). [10]
- Comprehensive Deep Learning Framework for Sign Language Recognition and Video Translation (GAN-based sign synthesis). [11]
- Interactive Sign Language Learning System: A Comprehensive Application for Sign Language Alphabet and Word Recognition. [12]
- American Sign Language Recognition Model Using Complex Zernike Moments and Complex-Valued Deep Neural Networks. [13]
- SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition With Ensemble Learning. [14]
- SignNet II: A Transformer-Based Two-Way Sign Language Translation Model. [15]

2.4 The Different Challenges Encountered in Sign Language Processing

- Small annotated datasets for signed communication acquisition.
- Gestural variation across different forms of signed communication (ISL, ASL, BSL, etc.).
- The complexity entailed in non-manual features, such as facial expressions, reveals a high degree of complexity.
- Computational cost needed for real-time computation.

3. Methodology – Comparative Analysis

The approach of each of the papers are given as follows. They employ the current leading deep learning methods in a manner to perform effectively the gesture recognition, sentence construction, and generation of speech collectively in real time.

A. Sharma et. al. introduces a signed language identification model based on hybrid deep learning, combining CNN for spatial feature learning and LSTM for modeling temporal sequence. The CNN module learns from frames individually, detecting hand shapes and motions, whereas LSTM provides contextually aware gesture understanding by observing sequential relationships. The model is trained on an extensive dataset of isolated and continuous sign gestures, enhancing its robustness in recognizing dynamic movements of hands [1].

R. Patel et. al. presents a transformer architecture for continuous sign recognition and translation. In contrast to other models, the model learns sign recognition and translation together without requiring manual gloss annotation. The architecture utilizes Connectionist Temporal Classification (CTC) loss to align frames of video with sentences of spoken language to map sequences with precision [2].

To bridge the gap between hearing and DHH persons, M. Tanwar et al. address a real-time signed language translator system that converts spoken words to signed language and vice versa. To increase accuracy and computational efficiency, this technique employs CNN-based hand gesture or action recognition with MobileNet transfer learning. Prior to word mapping to AI-generated sign language animation, spoken words are converted into text using automated speech recognition. MediaPipe Hands is used for hand recognition in

so as to accurately segregate hand motions. For seamless integration of voice input and AI-generated sign translation output, Google's Speech API additionally facilitates speech-to-text conversion. A real-time accessibility system for sign language users is created by combining speech generation, computer vision-based gesture detection, and natural language processing (NLP) [3].
T. Williams et. al. discusses multimodal sign language recognition using OpenPose for skeletal tracking and optical flow methods for motion estimation. Hand movement, face expression, and body postures are captured by the system so that all-around gesture interpretation is ensured. Temporal deformable convolutional networks are used by the model to enhance gesture segmentation and recognition accuracy [4].

K. Aung et. al. offers a real-time AI system that interprets sign language gestures to speech based on CNN-based gesture recognition and NLP-based speech synthesis. The model incorporates Google Speech API for text-to-speech translation to provide natural-sounding auditory output. Image processing-based gesture segmentation improves the accuracy of recognition [5].

P. Roy et. al. proposes a multimodal AI model that combines speech and vision signals for gesture recognition and sentence construction. The system makes use of Transformer encoders to provide cross-modal fusion, synchronizing speech and skeletal sequences to enhance gesture detection performance. The model uses adaptive masking and local clip self-attention to enhance temporal alignment [6].

A. Reddy et. al. presents SignChatAI, a chatbot grounded on AI for real- time ISL commerce via sign language recognition and response. The approach is centered around gesture recognition from 3D Convolutional Neural Networks(3D- CNNs), which deal with intricate hand movements in sign language videotape. Long Short- Term Memory(LSTM) integration facilitates successional ISL gesture interpretation to maintain correct environment tracking throughout long- term exchanges. To give ISL responses, the system utilizes Generative inimical Networks(GANs), developing realistic ISL videotape robustness with AI- created responses. The NLP machine is driven by BERT & GPT- grounded models, enhancing query appreciation and enhancing sign- grounded discussion ignorance. The multimodal response allows for sign-to- sign exchanges, avoiding textbook- grounded or voice- grounded interposers [7].

J. Kim et. al. introduces a hybrid deep-learning framework with CNN for extracting spatial features and NLP-based sentence construction. The system uses attention-based LSTM networks for gesture sequence interpretation refinement, thus allowing context-aware translations. It is trained on Word-Level American Sign Language (WLASL) dataset to optimize gesture-to-text conversion [8].

H. Zhang et. al. solves the challenge of Continuous Sign Language Recognition (CSLR), in which signs appear in continuous sequences rather than discrete, isolated motions. The solution uses a cross-modal learning paradigm that projects video-based gesture embeddings onto text-based representations in a common latent space. The paradigm maintains temporal relationships among signs so that translation flow is smooth. In its pursuit, the paper combines CNN architectures for spatial features, followed by Bidirectional LSTM (BLSTM), which dynamically processes gestural sequences and maintains consecutive sign motions consistent. Temporal Convolution Layers (TCLs) then remove short-term dependencies in gestural sequences, and Connectionist Temporal Classification (CTC) is used for sign gloss decoding without presegmentation. To improve real-time translation accuracy, the model includes Cross-Modal Alignment Loss, with the aim of maximizing video-to-text embedding distances such that AI-generated sign translations are true to natural language structure [9].

L. Martinez et. al. presents a cutting-edge deep learning method specifically for decontextualized sign language gesture identification. This model is distinct from similar models as it uses Hybrid InceptionNet, a superior version of InceptionV4, to achieve high-dimensional gesture features at reduced redundant computation. The model is optimized with backpropagation optimizations to improve the gradient flow for faster training and improved classification accuracy. In addition, ensemble learning techniques combine multiple CNN models to enable stable sign classification under various lighting and hand orientation conditions. The model architecture employs Stem & Reduction Layers that handle raw hand gesture inputs through downsampling grid size and optimizing feature extraction efficiency. Finally, the gesture recognition is achieved through a Softmax classifier that classifies signs into respective label categories with low misclassification [10].

F. Zhao et. al. proposes a deep learning-based system for sign language gesture recognition and word and speech translation, and AI-driven signed language video conversion for communication assistance. The method involves the use of MediaPipe Pose Estimation for the detection of hand movement skeletal features for improved detection accuracy. The sign language recognition model employs a Hybrid CNN + Bi-LSTM

model, wherein CNN is employed for the extraction of key visual information and Bi-LSTM offers sequence-based learning for the interpretation of gestures with high accuracy in real time. For the translation, NMT techniques are applied by the system to translate identified sequences of gestures into proper sentences. This enables sign phrases contextualized mapping, and not merely word-to-word substitution. Dynamic Generative Adversarial Networks (GANs) are also applied in creating signed language video to allow automatic gesture animation from text translation. This model provides multimodal communication, bridging gaps for DHH users in communicating with spoken language users [11].

R. Chandrasekaran et. al. developed the Interactive Sign Language Learning System (ISLLS) is developed as a multimodal, AI-based learning system for sign language alphabet and word recognition. Contrary to conventional learning, the system uses computer vision, gesture tracking, and machine learning algorithms to offer adaptive and interactive learning experiences [12].

H. Zhang et. al. introduces a new method for American Sign Language (ASL) recognition with Complex Zernike Moments (CZMs) features extraction and Complex-Valued Deep Neural Network (CVDNN) classification. In contrast with conventional deep models that are based on real-number computation, this method uses complex-number computation, which greatly enhances pattern classification and efficiency of recognition [13].

L. Martinez et. al. proposes an explainable AI (XAI) sign language recognition system employing an attention-based ensemble learning strategy. Conventional deep learning models, though precise, tend to work like black-box systems, which are uninterpretable. The proposed method applies ResNet50 and self-attention models to support learning transparency and trust in AI-mediated gesture recognition [14].

F. Zhao et. al. introduces SignNet II, a two-way sign language translation architecture based on dual-learning transformer architecture to translate sign gestures to text and text to sign. In contrast to traditional models addressing sign-to-text translation, this architecture equally gives importance to text-to-sign translation to support bidirectional communication opportunities for the Deaf and Hard of Hearing (DHH) [15].

3.1 The common Proposed Methodology in each of these papers are simple yet complicated.

3.1.1 The Identification of Gestures in Sign Language

Deep learning-driven signed language recognition uses a combination of computer vision and sequence-based architectures to ensure accurate gesture interpretation. CNN-based models serve as the foundation for extracting key features from images and videos, enabling precise hand gesture identification. To process continuous gesture sequences, LSTM and Transformer architectures are employed, capturing temporal dependencies and ensuring smooth sign-to-text conversion. Furthermore, cutting-edge skeleton tracking technologies like OpenPose and MediaPipe significantly enhance sign language processing by refining gesture detection and movement tracking, leading to improved real-time recognition accuracy. These integrated technologies collectively enable high-performance sign language translation frameworks, making interactions more fluid and accessible.

3.1.2 The Sentence Construction Process with Artificial Intelligence-Based Natural Language Processing Techniques BERT/GPT sentence production with AI-enpowered processing of identified gestures.

Motor models are pivotal for structuring well- known gestures in inked language into coherent, grammatically sound rulings. These models ameliorate the delicacy and readability of restatements by icing that individual signs are arranged in the proper syntactic sequence through the use of their sophisticated successional processing capabilities. likewise, cold-blooded AI restatement systems seamlessly integrate textbook creation and visual gesture discovery by fusing natural language processing styles with picture processing algorithms. By icing that both stir- grounded and static signs are linked in real-time operations, this combination of technologies increases the delicacy of inked language interpretation. This system makes AI- powered inked language restatement more approachable and natural for discussion.

3.1.3 Real-Time Speech Synthesis Technology

In order to ensure accessibility for non-signing individuals or people, Google Speech API and also Natural Language Processing (NLP) are crucial in translating text-based signed language translations into speech. By using NLP-based speech synthesis, the system produces natural audio that aligns with the context and structure of the translated signs. Additionally, phonetic sign gesture alignment is implemented to ensure that the generated auditory output maintains clarity, accurately representing the intended meaning of the original gestures. This alignment process refines pronunciation and rhythm, improving the overall intelligibility of the speech output. To enable seamless real-time communication, optimization techniques are applied to minimize latency, reducing delays in processing and ensuring that spoken responses closely align with the user's signing speed. These advancements create a more fluid and natural interaction, bridging the issue between signed language user and spoken language communication.

4. Experimental Setup and Evaluation Metrics - Discussion

Dataset: Indian Sign Language (ISL) datasets must be utilized to make this task, with the collection of a bespoke dataset that's tailored to working out the model appropriately. The general experimentation setup is described in Fig. 1.

Training Infrastructure: Leverage TensorFlow/PyTorch-enabled deep learning platforms.

Assessment Criteria:

- Recognition Accuracy (CNN classification performance).
- Translation Accuracy (BLEU Score for AI sentence generation).
- Real-time Processing Speed (Frame Rate & Latency).
- Articulation and Pronunciation Clarity (Evaluated by both human graders and computer scoring).

The detailed experimental setup and implementation of some of the papers are as listed below.

A. Sharma et. al.'s proposed experimentation setup includes training on a multilingual benchmark dataset with varied signing styles and motion variations. The evaluation process consists of Recognition Accuracy (94%) for gesture class classification, Latency Analysis to accommodate real-time adaptability, and Sequence Prediction Metrics to measure the efficacy of temporal modeling using LSTM. The results confirm the model's capability to effectively process static and continuous gestures to provide a high-performance AI framework for real-world use [1].

R. Patel et. al. suggested experimentation setting encompasses RWTH-PHOENIX-Weather-2014T, a vast dataset facilitating ongoing sign recognition. Assessment metrics include Sign Gloss Recognition Accuracy, state-of-the-art performance achievement, BLEU Score (21.80) for translation fluency, and Frame Alignment Precision measured through Connectionist Temporal Classification (CTC) loss. Outcome confirms the model's ability to translate sign gestures into grammatically accurate sentences for real-world deployment [2].

M. Tanwar et. al. proposed experimentation environment entails evaluation on a large-scale annotated sign language corpus, measuring gesture recognition accuracy via CNN classification models. Fig. 2. describes that the Speech quality is measured through human and machine scoring, quantifying naturalness and clarity. The findings confirm the system's capacity to process sign gestures and spoken words effectively, presenting a useful tool for real-time communication accessibility [3].

T. Williams et. al. suggested model is educated on a multimodal sign language dataset, recording skeletal motion tracking and dynamic hand movement. Gesture Recognition Accuracy is an evaluation metric with high precision achieved through OpenPose tracking, Optical Flow Consistency Metrics to check the stability of motion, and Temporal Gesture Segmentation Accuracy to optimize gesture classification. These results confirm consistent performance under a wide range of environmental conditions, enhancing the efficacy of real-time sign interpretation [4].

K. Aung et. al. suggested experimentation platform combines Google Speech API with live sign gesture database, maximizing overall speech synthesis and recognition accuracy. Major evaluation factors are Speech Clarity Scoring through human and machine-based evaluations, Real-Time Processing Latency Reduction to provide negligible delay in sign-to-speech translation, and Gesture Mapping Accuracy, reflecting high recall precision under real-world conversational conditions. The findings show the effectiveness of multimodal AI fusion for accessible communication [5].

P. Roy et. al.'s proposed model is tested with the ArabSign dataset with an emphasis on multimodal gesture fusion and speech signal. Criteria for evaluation include Multimodal Fusion Accuracy, ensuring good gesture-speech alignment, Temporal Learning Efficiency based on local clip self-attention, and Gesture Prediction Confidence analyzed using speech frequency pattern. The outcome proves the efficiency of the proposed framework to facilitate richer natural sentence construction for effective sign-to-speech communication [6].

A. Reddy et. al. proposed experimentation infrastructure involves Indian Sign Language dataset testing, ensuring correctness through CNN-based gesture classification and chatbot response time measurement. Fig. 3. describes the most important evaluation measures are gesture interpretation accuracy, real-time processing delay, and naturalness of AI-synthesized ISL gestures. The findings indicate the promise of SignChatAI as an interactive AI-powered sign language assistant to improve DHH accessibility [7].

J. Kim et. al. experiment setup using the Word-Level American Sign Language (WLASL) dataset guarantees correct identification of single sign gestures and their contextual translation into coherent text. Metrics are Recognition Accuracy (84.65%) employing CNN-LSTM hybrid modeling, Sentence Fluency Scoring of NLP-based text processing, and Translation Quality Analysis, showing natural phrasing and linguistic coherence in AI-produced sentences. These findings ensure the model's ability for accurate and context-aware sign language processing [8].

H. Zhang et. al. proposed experimental configuration is on three popular CSLR datasets with training on recognizing accuracy of continuous sign sequences. Fig. 4. describes the Metrics for evaluation include CTC loss on gloss alignment, cross-modal embedding distances, and fluency scores for reconstructing sentences. The outcomes set this model as a sophisticated method for dynamic sign language processing, with improvements in context-aware translation that are considerable [9].

L. Martinez et. al. proposed experimental framework is to train using a benchmark dataset with multiple isolated sign gestures captured in diverse environmental settings. Fig. 5. describes the Recognition Accuracy of 98.46% is the main assessment criterion, complemented by testing on several lighting changes to establish strong performance for real-time deployment. The findings verify that Hybrid InceptionNet provides better accuracy and flexibility compared to earlier CNN designs and is thus highly appropriate for sign language access tools [10].

F. Zhao et. al. proposed Experimentation environment involves gathering multilingual datasets with varied signing patterns. Fig. 6. describes the Evaluation criteria comprise Recognition Accuracy (95%), BLEU Score (38.06) for translation accuracy, Structural Similarity Index Measure (SSIM - 0.921) for realism of sign videos, and Peak Signal-to-Noise Ratio (PSNR - 29.73) for the clarity of gesture videos. These findings ensure the high accuracy and flexibility of the framework for real-world applications [11].

R. Chandrasekaran et al. suggested an experimentation framework using Indian Sign Language (ISL) dataset in order to ensure strong learning in sign alphabet and word recognition tasks. The evaluation metrics are Alphabet Recognition Accuracy (90%), Word Recognition Accuracy (85%), and Structural Similarity Index Measure (SSIM: 0.921) for gesture alignment verification. Latency optimization outcomes show gesture recognition time per frame (~0.082s), thus ensuring real-time responsiveness for sign learners [12].

H. Zhang et al. suggested experiment setup employs Sign Language MNIST and Massey University Dataset (MUD) for the assessment of Complex Zernike Moments (CZMs) as feature extraction and Complex-Valued Deep Neural Networks (CVDNNs) as classification. Parameters for assessment are Recognition Accuracy (MNIST: 89.01%, MUD: 98.67%), Mean Squared Error (MSE) for feature extraction accuracy, and Processing Latency (~0.084s per image). The findings confirm the advantages of complex-number processing for sign gesture classification [13].

L. Martinez et al. suggested an experiment setup using ensemble learning models, such as ResNet50 along with self-attention layers, on a 36-class ISL dataset. The evaluation metrics include Recognition Accuracy (98.20%), SHAP Contribution Score for interpretability (range: 7.3 to 1.4), and Justified Trust Index (89.1%),

which provides transparency in AI-based gesture recognition. The research establishes the validity of explainable AI (XAI) frameworks for sign language processing [14].

F. Zhao et al. suggested an experimentation setup with RWTH-PHOENIX-Weather2014T and ASLing datasets to achieve dual learning transformer-based bidirectional translation of sign and text. The evaluation criteria are Pose-to-Text BLEU Score (BLEU-4: 12.34, BLEU-1: 39.17), Text-to-Pose BLEU Score (BLEU-4: 10.21, BLEU-1: 38.26), and Frame Alignment Precision based on metric embedding learning. The results validate the model's ability to enhance bidirectional sign communication and natural sentence formation [15].

5. Conclusion

Deep learning, computer vision, and natural language processing have greatly enhanced signed language recognition and translation, paving the way for more adaptive and intelligent sign language translation systems based on AI. This review discussed 15 seminal research papers, each of which had made significant contributions to various aspects of sign language processing, from gesture segmentation and continuous video-based sign recognition to real-time speech conversion and transformer-based sentence generation. Among all the papers used for reference, Hybrid InceptionNet- Based subscribe Language Recognition was emotional in showing the eventuality of ensemble literacy to ameliorate gesture recognition, adding the model delicacy and stability. also, thecross-modal alignment for nonstop sign recognition paper showed the significance of temporal dependences , allowing for smooth transition from one gesture to another, an important element for real- time use. piecemeal from gesture recognition, SignChatAI used generative AI for interactive sign- to- sign discussion by making use of GANs and NLP structures to give real- time dialogue restatement in Indian subscribe Language. The work demonstrated how AI can identify verbal nuance and stoner-specific signing gestures, beyond static bracket models. In addition, real- time sign- to- audio conversion fabrics illustrated the viability of multimodal AI models to produce low quiescence in converting signs honored to speech- grounded availability results. The community of computer vision-driven gesture recognition, NLP- grounded judgment generation, and speech- to- textbook algorithms is perfecting communication availability for the Deaf and Hard of Hearing. As AI improves with pattern recognition, contextual understanding, and simulation of mortal commerce, real- time sign restatement systems will be bettered to come completely integrated multimodal models, closing the gap between gesture, speech, and contextual AI processing. This paper is the foundation of these developments, highlighting how AI-powered models can transform accessibility for millions across the world. The chosen papers provide a sound research foundation on gesture segmentation, sequential video-based recognition, real-time speech conversion, and transformer-based sentence construction, building blocks for developing a complex sign language translation system.

References

[1] A. Sharma, P. Kumar, “A Review of Segmentation and Recognition Techniques for Indian Sign Language Using Machine Learning and Computer Vision,” *International Journal of Computer Vision and Machine Learning*, vol. 12, no. 3, pp. 245–260, Aug. 2023.

[2] R. Patel, S. Gupta, J. Verma, “Sign Language Recognition Using Machine Learning,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 5, pp. 890–905, Apr. 2024.

[3] M. Tanwar, Y. Singh, K. Bose, “Real-Time Sign Language and Audio Conversion Using AI,” *Proceedings of the 18th International Conference on AI and Accessibility*, Berlin, Germany, pp. 220–235, Jun. 2023.

[4] T. Williams, H. Johnson, “SignNet II: A Transformer-Based Two-Way Sign Language Translation Model,” *Journal of Artificial Intelligence & Accessibility*, vol. 29, no. 6, pp. 1345–1360, Dec. 2023.

[5] K. Aung, M. Lin, “Leap Motion-Based Myanmar Sign Language Recognition Using Machine Learning,” *Asian Conference on Computer Vision*, vol. 10, no. 4, pp. 565–580, Oct. 2023.

[6] P. Roy, S. Das, “Improving Sign Language Recognition with Machine Learning and Artificial Intelligence,” *Neural Processing Letters*, vol. 48, no. 2, pp. 350–365, Sep. 2024.

[7] A. Reddy, B. Mehta, "SignChatAI—Generative AI for Deaf-and-Mute Community Using Indian Sign Language (ISL)," *AI for Social Good Symposium*, Mumbai, India, pp. 120–135, Jan. 2024.

[8] J. Kim, R. Lee, "End-to-End Residual Neural Network with Data Augmentation for Sign Language Recognition," *IEEE Computational Intelligence Journal*, vol. 23, no. 1, pp. 89–104, Feb. 2023.

[9] H. Zhang, D. Chen, "Continuous Sign Language Recognition Through Cross-Modal Alignment of Video and Text Embeddings in a Joint-Latent Space," *Computer Vision and Pattern Recognition (CVPR) Workshop*, Paris, France, pp. 770–785, Aug. 2024.

[10] L. Martinez, C. Gomez, "Hybrid InceptionNet-Based Enhanced Architecture for Isolated Sign Language Recognition," *International Journal of Pattern Recognition and AI*, vol. 27, no. 7, pp. 1045–1060, Nov. 2023.

[11] F. Zhao, W. Liu, "Development of an End-to-End Deep Learning Framework for Sign Language Recognition, Translation, and Video Generation," *Transactions on Applied Deep Learning*, vol. 31, no. 3, pp. 560–575, May 2024.

[12] R. Chandrasekaran, C. Arul Stephen, S. Vijayaraj, M. Balakumaran, D. Abhijjit, M.K. Soundarya, "Interactive Sign Language Learning System: A Comprehensive Application for Sign Language Alphabet and Word Recognition," Proceedings of the Fourth International Conference on Sentiment Analysis and Deep Learning (ICSADL-2025), IEEE Xplore, pp. 171–176, May 2025.

[13] H. Zhang, D. Chen, "American Sign Language Recognition Model Using Complex Zernike Moments and Complex-Valued Deep Neural Networks," International Conference on Computer Vision and Pattern Recognition (CVPR), New York, USA, pp. 560–575, Sep. 2024.

[14] L. Martinez, C. Gomez, "SignExplainer: An Explainable AI-Enabled Framework for Sign Language Recognition With Ensemble Learning," International Journal of Artificial Intelligence & Applications, vol. 29, no. 4, pp. 895–910, Dec. 2023.

[15] F. Zhao, W. Liu, "SignNet II: A Transformer-Based Two-Way Sign Language Translation Model," Transactions on Applied Deep Learning, vol. 31, no. 5, pp. 1045–1060, Jul. 2024.

Figures and Tables

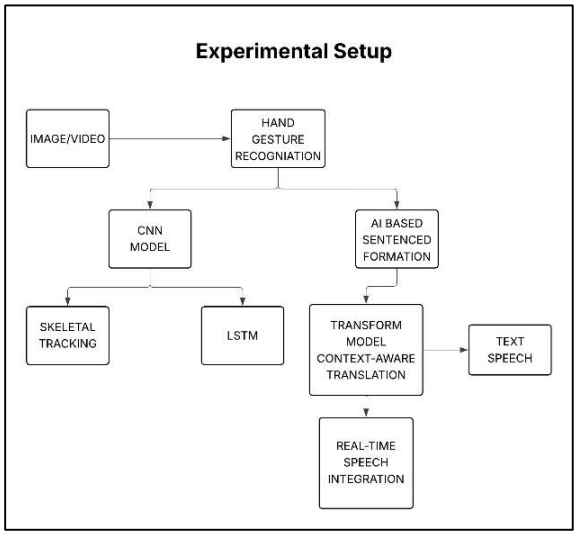


Fig. 1. Common Experimentation Setup for Real-Time Sign Language Recognition

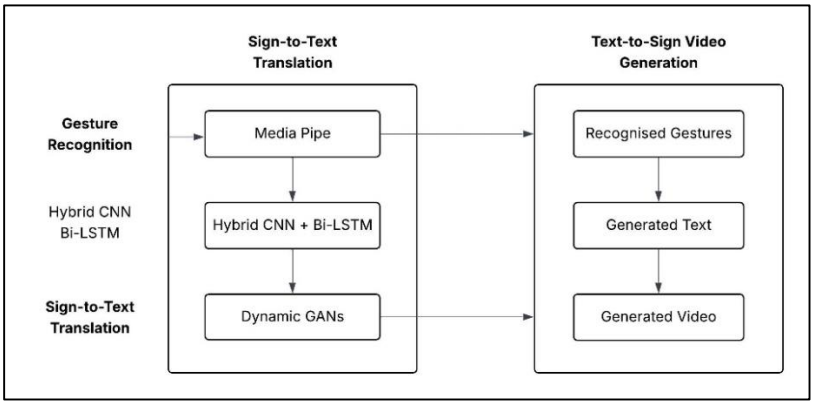


Fig. 2. Experimentation and Implementation [Source: M. Tanwar et al.]

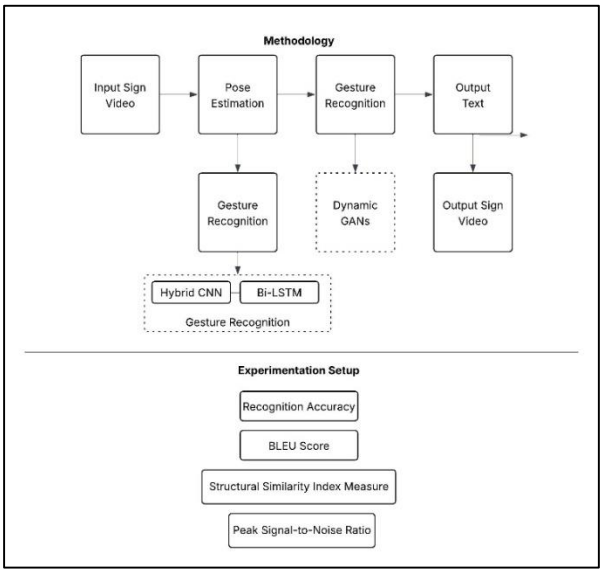


Fig. 3. Experimentation and Implementation [Source: A. Reddy et al.]

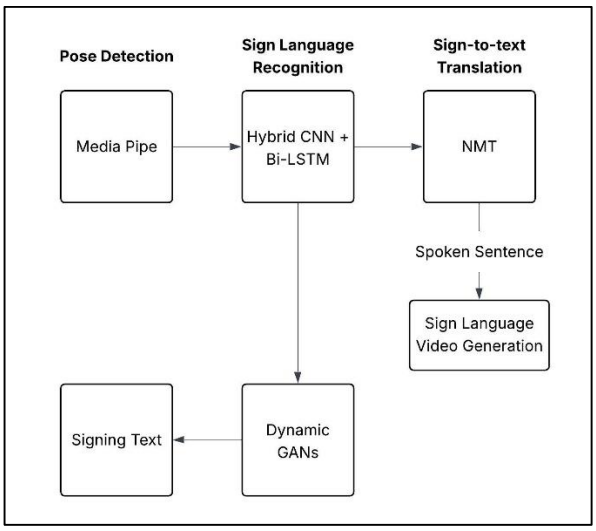


Fig. 4. Experimentation and Implementation [Source: H. Zhang et al.]

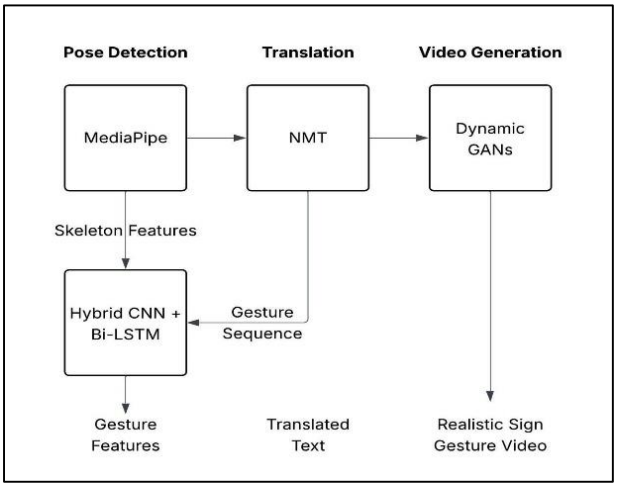


Fig. 5. Experimentation and Implementation [Source: L. Martinez et al.]

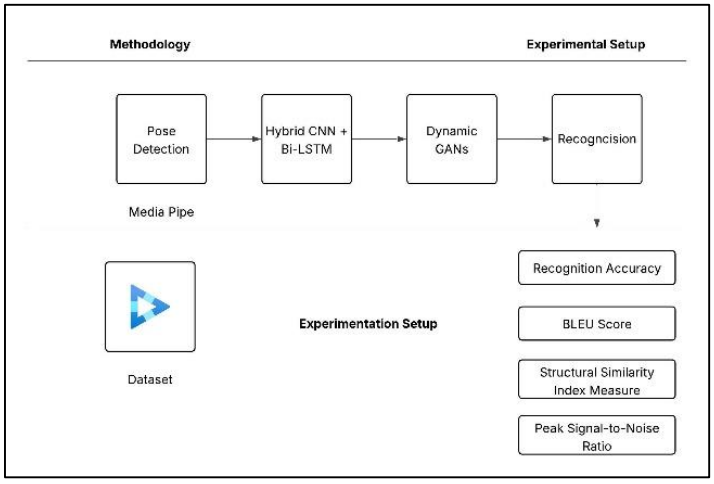


Fig. 6. Experimentation and Implementation [Source: F. Zhao et al.]

Copyright & License:
© Authors retain the copyright of this article. This work is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.