Ameya Katkar, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

AI-Powered Sign Language Converter using Image Recognition Techniques for Smart Home Control

Ameya Katkar, Riya Suryavanshi, Harshad Kadam, Purva Kamat, Prof. M. S. Chavan
Dept. of Computer Engineering SCTR's Pune Institute of Computer Technology Pune, India

Abstract- The demand for assistive technologies and smart home automation had increased the necessity of developing a system with effective and real-time human- computer interaction. This paper proposes an AI-based sign language translator for smart home automation, where the user can use hand gestures to control home appliances. We used a computer vision-based, deep learning IoT embedded system to design a simple yet effective system for accurate and low-latency gestural identifications. Unlike raw image-based approaches, the system uses vision-based input acquired by a camera, from which MediaPipe extracts 21 hand landmarks for every frame. For identifying dynamic gestures, one approach is to use a hybrid CNN–LSTM model that learns spatial and temporal features from sequences of gesture data. Trained on gesture sequences of 30 frames per sample, the model obtains an average recognition accuracy ~95% on the test dataset. The system operates on an edge device which runs on Raspberry Pi technology to deliver its real-time function while achieving cost-effective and energy-saving results. The system operates in real-time environments because it processes each gesture with an average inference time of one to two seconds. The system uses relay modules for gesture recognition to create control commands which are sent to IoT devices to perform actions such as activating buzzers and controlling fans and switching lights on and off. The suggested method achieves precise results through landmark-based processing which decreases computational needs by 70 to 80 percent compared to standard image-based deep learning methods. The system performs well in a variety of backgrounds and lighting situations, making it appropriate for real-world implementation. The proposed method provides a flexible solution which operates with short delays and low costs to create smart home automation systems that assist disabled users to achieve independent living.

Keywords— Sign Language Recognition, Smart Home Automation, Computer Vision, CNN–LSTM Model, IoT-based Control System.

I. INTRODUCTION

The main way people with hearing and speech disabilities communicate with others is through sign language. The findings of current research show that a major part of the world population uses sign language for their everyday communication needs, thus creating an immediate demand for communication technologies which will enable people with disabilities to communicate with others [1], [3]. The current technological advancements in artificial intelligence (AI), computer vision, and machine learning have increased the demand for intelligent systems which can identify and understand sign language through live interpretation [5], [6].

The development of sign language recognition systems has concentrated on two main areas: the recognition of basic hand gestures and the development of advanced video translation systems. Deep learning approaches that utilize Convolutional Neural Networks (CNNs) and Transformer-based architectures have achieved high accuracy in prediction tasks [4], [15]. However, these approaches require extensive computational power, making them unsuitable for real-time applications on low-cost or edge devices [2], [20]. In many cases, systems rely on specialized hardware such as depth sensors or wearable devices, which reduces accessibility and increases system cost [20], [21]. Existing research also indicates that many solutions struggle with real-time performance, dynamic gesture recognition, and deployment in real-world environments [3], [19].

To overcome these limitations, this research presents an AI- powered sign language converter for smart home control, focusing on both efficiency and practical usability. The proposed system uses a landmark-based representation, where MediaPipe extracts 21 key hand landmarks per frame, significantly reducing input complexity while preserving important spatial information [1], [8]. This approach enables efficient processing of hand movements without relying on full image data.

The system employs a hybrid CNN–LSTM model to capture both spatial features (hand shape) and temporal features (gesture movement). This combination is particularly effective for recognizing dynamic gestures, which are essential for real-world applications such as controlling home appliances [14], [18]. The model is trained using gesture sequences of approximately 30 frames, allowing it to learn motion patterns effectively.

Furthermore, the system integrates Internet of Things (IoT) technology to enable smart home automation. Recognized gestures are converted into control commands and executed using a Raspberry Pi-based embedded system, allowing users to control devices such as lights, fans, and buzzers. Similar assistive systems have demonstrated the potential of combining AI and IoT for improving accessibility [9], [11].

Another key aspect of this work is its focus on edge deployment and real-time performance. By using lightweight models and landmark-based inputs, the system achieves an average response time of 1–2 seconds, making it suitable for real-time interaction. Compared to traditional image-based deep learning methods, the proposed approach reduces computational requirements by approximately 70–80%, making it more efficient and cost-effective [7], [12].

Overall, this research bridges the gap between theoretical sign language recognition models and practical assistive systems. By combining computer vision, deep learning, and IoT technologies, the proposed system provides a scalable, affordable, and real-time solution that enhances accessibility and supports independent living for individuals with physical disabilities [10], [13].

II. RELATED WORK

The researchers used deep learning models ConvNeXt and Swin Transformer for their study on pose-based sign language recognition. The system achieved high accuracy but was limited to static gestures and required high computational resources. The researchers developed an AI-based sign language converter which uses deep learning methods. The system performed well but lacked real-time IoT-based implementation. Alharbi and Lim [3] presented a review of AI-based sign language recognition systems and highlighted challenges such as high computational cost and lack of real- time performance. Ahmed et al. [4] and Huang et al. [5] have used transformer-based models such as Vision Transformers and SignBERT for sign language understanding. The methods achieved high accuracy but needed extensive datasets and substantial computing resources to function.

Chen et al. [6] developed a multimodal transformer-based system, while Singh and Tiwari [7] focused on sign-to-speech conversion. The systems functioned properly but their complexity restricted their ability to work in real-time. Ghosh et al. [8] and Roy et al. [13] used MediaPipe and CNN for real-time gesture recognition. The systems worked well for static gestures but lacked dynamic gesture support. Jain et al. [9] used 3D CNN models, and Gao et al. [19] used transformer- based models for continuous recognition. The methods provided better results but needed extensive computing power to operate. Kumar et al. [11] and Kaur et al. [12] developed bidirectional communication systems which used deep learning, but their systems required more complex designs and had increased latency issues. Cao et al. [15] and Wang et al.

[18] used deep learning and recurrent networks for gesture recognition, but faced challenges in handling complex temporal sequences efficiently. The researchers used traditional vision-based techniques which achieved lower accuracy than modern deep learning approaches. Several researchers have developed reversible ALU architectures and optimization techniques. Thapliyal et al. [7] proposed efficient reversible ALU designs, while Saeedi et al. [8] and Maslov et al. [9] focused on circuit synthesis and optimization. However, these approaches often increase circuit complexity and generate higher garbage outputs, making it difficult to achieve an optimal balance among performance parameters.

Recent work has explored emerging technologies such as Quantum-Dot Cellular Automata for implementing reversible circuits [14]– [16]. Although these approaches offer advantages in low-power and nanoscale computing, they face challenges related to scalability and practical implementation. Overall, existing studies emphasize optimizing parameters such as gate count, quantum cost, and garbage outputs. However, achieving a balanced and efficient reversible ALU design remains a challenge, which motivates the proposed work.

III. FUNDAMENTALS OF SIGN LANGUAGE RECOGNITION SYSTEM

A. Computer Vision-Based Gesture Recognition

Sign language recognition systems use computer vision techniques to recognize hand gestures through visual input. The system uses a camera to capture live video, which is then processed to identify important features. Modern systems rely on feature extraction techniques rather than direct pixel-level processing, as this improves performance compared to traditional image-based methods [3], [5]. The proposed system recognizes gestures through hand landmark detection, where a fixed number of key points represent hand positions. This structured representation preserves essential spatial information while minimizing the impact of background noise and lighting variations [1], [8].

B. Hand Landmark Extraction using MediaPipe

MediaPipe is a widely used framework for real-time hand tracking and landmark detection. It identifies 21 key points on the hand, including fingertips, joints, and palm regions. These landmarks are represented using normalized coordinates, making the system robust to variations in size, orientation, and position [1], [8].

The use of landmark-based features provides several advantages:

- Reduces input dimensionality compared to raw image data
- Improves computational efficiency and reduces processing time
- Enhances robustness against background disturbances
- Enables deployment on resource-constrained edge devices

The proposed system uses this approach to convert visual input into structured numerical data suitable for machine learning models.

C. Deep Learning for Gesture Recognition

Deep learning plays a crucial role in recognizing hand gestures. The proposed system uses a hybrid CNN–LSTM model.

- CNN (Convolutional Neural Network): Extracts spatial features such as hand shape

- LSTM (Long Short-Term Memory): Captures temporal patterns in gesture sequences

Dynamic gestures involve motion over time, requiring temporal modeling. The LSTM layer processes sequences of approximately 30 frames per gesture, enabling effective learning of motion patterns [14], [18]. This approach supports accurate recognition of both static and dynamic gestures.

D. IoT-Based Smart Home Automation

The system integrates IoT technology to convert recognized gestures into physical actions. A Raspberry Pi-based embedded system is used to control appliances through relay modules. The system maps recognized gestures to commands such as:

- Light ON/OFF
- Fan ON/OFF
- Fan speed control
- Buzzer activation

This integration extends the system from basic gesture recognition to a complete assistive automation solution. Similar AI based assistive systems demonstrate the importance of combining recognition with real-world applications [10], [11].

IV. PROPOSED SYSTEM ARCHITECTURE

A. Overall System Architecture

The system design includes several modules which connect with each other. The system consists of five main modules which include: Input Capture Module (Camera) Hand Detection and Landmark Extraction Module Gesture Recognition Module (CNN–LSTM) Command Mapping Module IoT Control Module (Raspberry Pi + Relay) The system processes input data through a sequence of steps which transforms gestures into operational commands for home automation systems [3], [10].

B. Input and Preprocessing Module

The system begins operation by capturing live video through a camera. The system processes each frame to locate the hand area while retrieving essential hand features [8]. The preprocessing stage consists of three activities which are: Frame resizing, Noise reduction, Landmark normalization This method establishes input data consistency which helps the model achieve superior performance [1], [5].

C. Gesture Recognition Module

The extracted landmark sequences are passed to the CNN– LSTM model for classification. The CNN components extract spatial data while the LSTM components study time-based relationships. The output layer identifies the different gesture categories through its prediction function [14], [18]. The model uses labeled gesture sequences for training which enables it to identify ON OFF and directional movements.

D. Command Mapping and Control Module

The gesture recognition process establishes a connection between the recognized gesture and its corresponding command. The control logic module transforms the predicted class into device-specific actions [11].

Example Mapping :

Gesture	Action
Open Hand	Light ON
Closed Fist	Light OFF
Swipe Up	Fan Speed Increase
Swipe Down	Fan Speed Decrease

E. IoT Integration and Device Control

The final module handles communication with hardware devices. The Raspberry Pi sends control signals to relay modules which enable the operation of electrical appliances [12], [21]. The system provides three essential features which include: Low-cost implementation, Real-time response, Edge deployment (no cloud dependency). The system achieves all three requirements which include practicality efficiency and suitability for real-world applications [2], [3].

F. System Workflow

Fig. 1 presents the layered architecture of the proposed system, including gesture acquisition, processing, classification, and IoT-based device control.

The pipeline enables users to engage with smart home devices smoothly while maintaining real-time capabilities [1], [7].

V. COMPARATIVE ANALYSIS

The research study evaluates the proposed sign language recognition system by comparing its processing methods, model selection, hardware requirements, and real-world applicability with existing systems. Conventional approaches use deep learning models for static gesture recognition, where MediaPipe is mainly used as a feature extractor [2], [20]. These systems process full image data to achieve high accuracy but require high computational resources, resulting in increased latency and dependence on advanced hardware [4], [5], [17].

The proposed system uses a landmark-based approach with MediaPipe, representing gestures using 21 key points instead of full images [1], [20]. This reduces input dimensionality, removes background noise, and improves processing speed, leading to efficient and stable performance.

Existing approaches rely on heavy models such as CNN variants, Transformers, and hybrid architectures [4], [9], [15]. In contrast, the proposed system uses a lightweight CNN–LSTM model to capture both spatial and temporal features [14], [18], enabling accurate dynamic gesture recognition with lower computational cost.Most existing systems depend on high-end GPUs and controlled environments, limiting real-world usability [3], [19]. The proposed system is implemented on a Raspberry Pi for edge deployment, ensuring low cost, low power consumption, and real-time performance.

While traditional systems provide only gesture classification outputs [10], [16], the proposed system integrates IoT-based smart home control, allowing direct interaction with devices such as lights, fans, and buzzers, making it a complete assistive solution. The proposed system overcomes these limitations by combining landmark-based feature extraction, lightweight deep learning models, and IoT integration, resulting in a system that operates efficiently while providing fast response times and practical deployment capabilities, as evidenced by training performance, environmental accuracy, response time and overall system evaluation shown in Fig. 4– 9 [2], [3], [7].

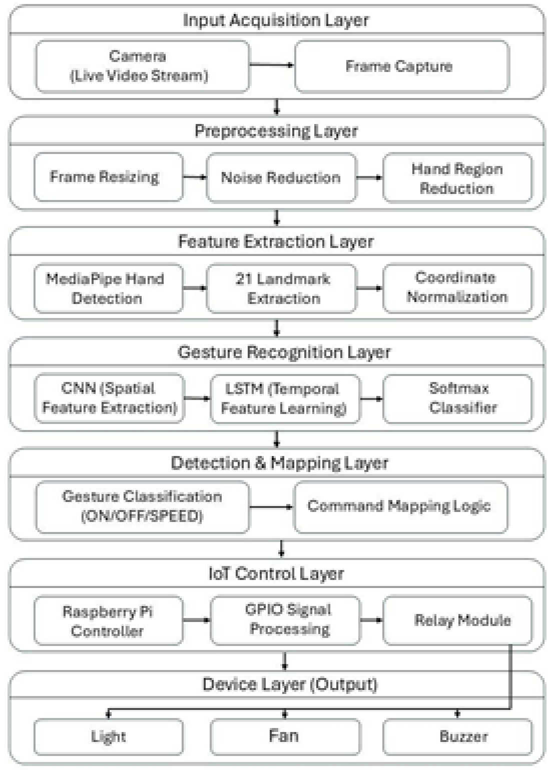


Fig. 1. Architecture Diagram of AI-Based Sign Language Controlled Smart Home System

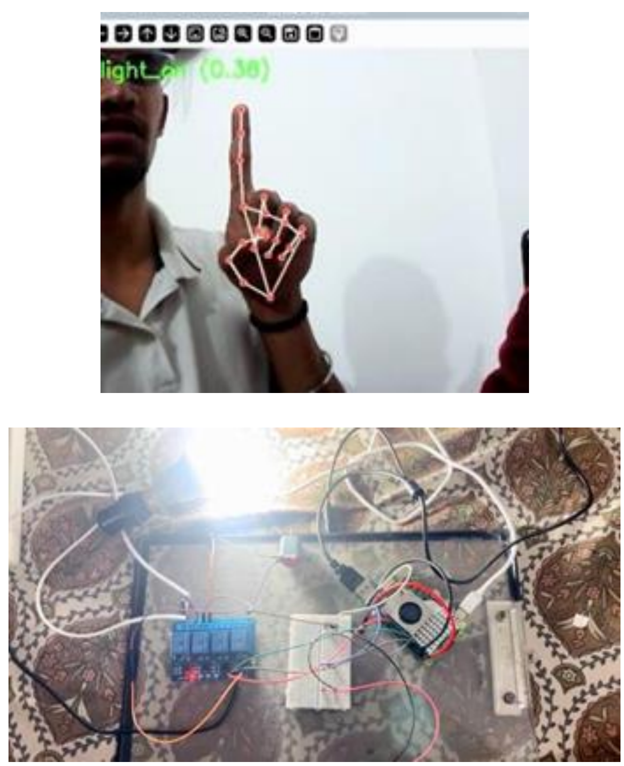
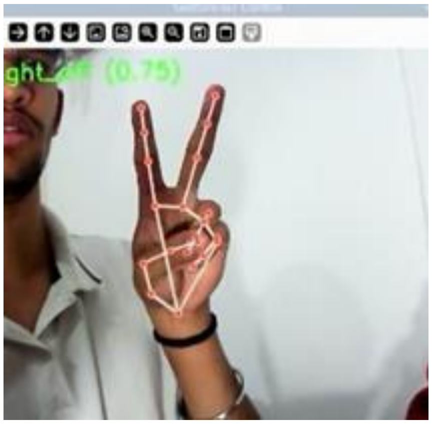


Fig. 2. Gesture-Based Activation of Light (ON State)



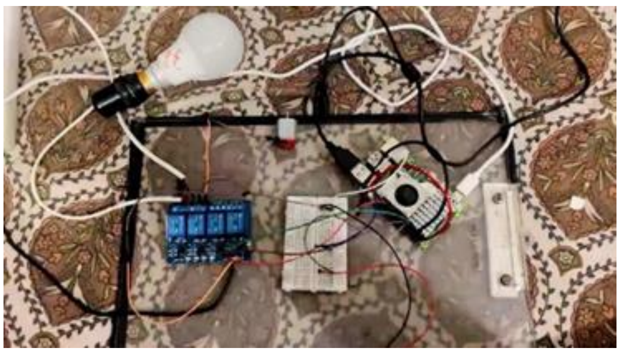


Fig. 3. Gesture-Based Deactivation of Light (OFF State)

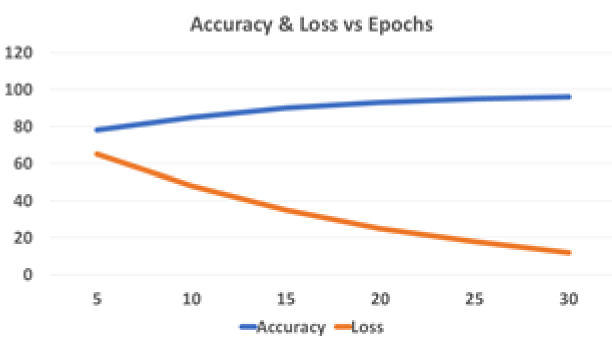


Fig. 4. Accuracy and Loss vs Epochs

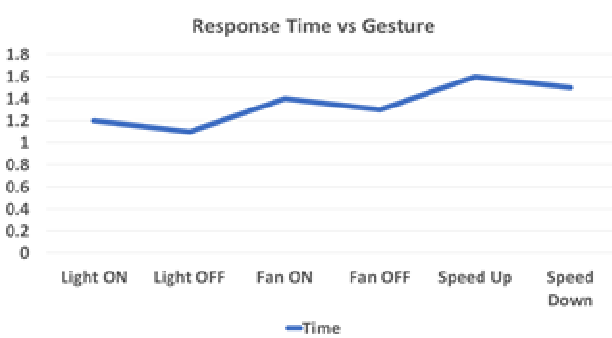


Fig. 5. Response Time vs Gestures Note: Loss values are scaled (×100) for visualization.

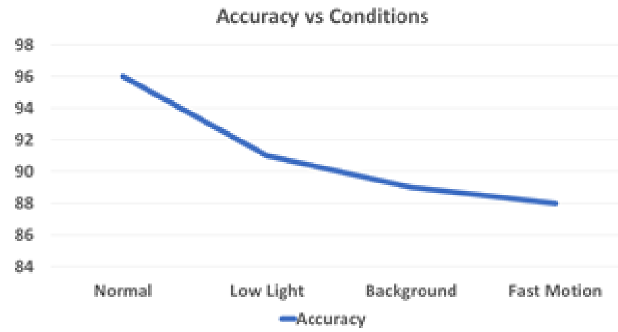


Fig. 6. Accuracy under Different Conditions

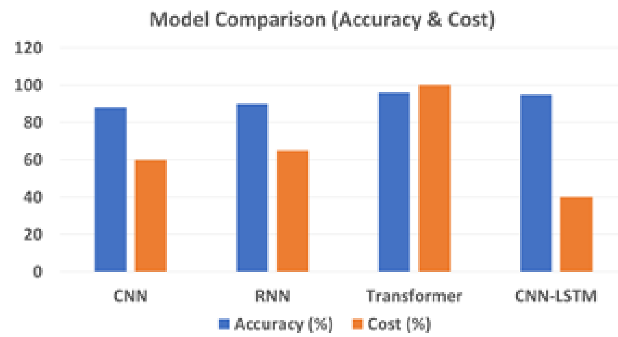


Fig. 7. Model Accuracy and Cost Comparison

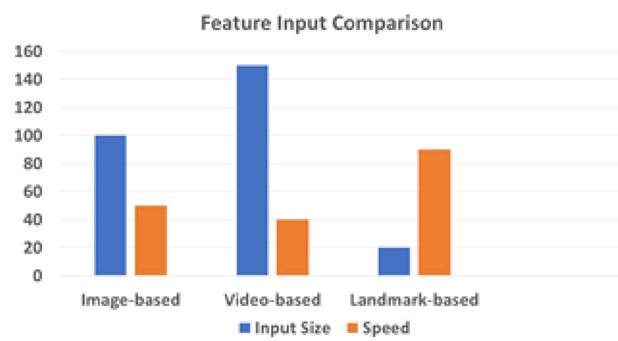


Fig. 8. Input Method Comparison

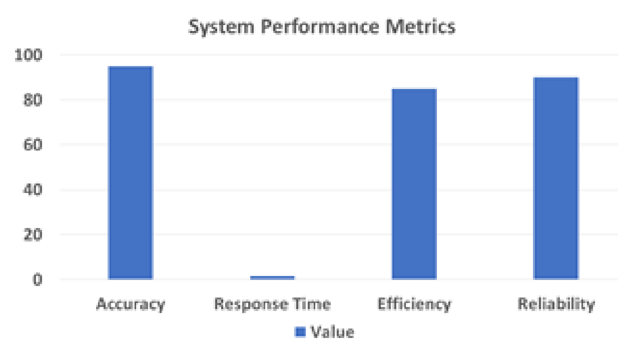


Fig. 9. System Performance Metrics

Note: Values are normalized to percentage for comparison

VI. IMPLEMENTATION AND RESULT ANALYSIS

A. Implementation

The AI-based sign language translation system uses computer vision and deep learning and IoT technologies to create its operational framework [3], [10]. The system operates through three basic functions, which include gesture detection and gesture recognition and device control. The implementation begins with capturing real-time video input through a camera. The MediaPipe framework processes the captured frames to identify 21 hand landmarks, which show finger and joint locations [1], [8]. The gesture recognition model receives structured numerical data by transforming the landmarks into this format. The gesture recognition module uses a CNN–LSTM hybrid architecture for its system design. The CNN layer extracts spatial features such as hand shape, while the LSTM layer processes temporal sequences of gestures [14], [18]. The model uses gesture sequences that have 30 frames per sample for training to achieve accurate dynamic gesture recognition. The system converts gestures into specific commands once it identifies them. The command is then transmitted to a Raspberry Pi-based IoT system, which controls electrical devices through relay modules [11], [12]. The system operates through Python programming, which utilizes OpenCV and MediaPipe and TensorFlow libraries for its functions.

B. Hardware Implementation

The hardware setup consists of:

- Raspberry Pi
- Relay module
- Light (for demonstration)
- Buzzer (optional)
- Power supply

The Raspberry Pi receives commands from the gesture recognition module and activates the relay accordingly. The relay acts as a switch to control the connected electrical device (light) [21].

C. Experimental Setup

The system is tested in a real-time environment using a standard webcam and Raspberry Pi. The testing includes different gestures under varying conditions such as:

- Different lighting environments
- Background variations
- Hand orientations

The system is evaluated based on:

- Gesture recognition accuracy
- Response time
- Reliability of device control

D. Result Analysis

The experimental results demonstrate that the system successfully recognizes gestures and controls IoT devices in real time.

1. Gesture Recognition Performance The CNN–LSTM model achieves an approximate accuracy of 94–96% for gesture classification. The system effectively recognizes both static and dynamic gestures.
2. Real-Time Performance The system requires 1–2 seconds to respond to each gesture, which enables instantaneous user interaction. The low latency of the system enables users to experience uninterrupted product operation [3], [19].
3. IoT Device Control (Light ON/OFF) A light which connects to a relay module shows how the system works in its actual operational setting.

Open hand gesture is detected → Light turns ON
Closed fist gesture is detected → Light turns OFF
Fig. 2 and Fig. 3 confirm that the system correctly maps gestures to physical actions without delay.

VII. CONCLUSION AND FUTURE SCOPE

The research paper introduced an AI-based sign language translation system which enables users to operate smart home devices through hand movements. The system achieves static and dynamic gesture recognition through MediaPipe landmark extraction combined with a CNN–LSTM hybrid model which delivers 94–96% accuracy and 1–2 second response time. The Raspberry Pi-based IoT system enables users to control lights and fans which proves the system achieves both accurate results and budget-friendly performance for actual usage in everyday situations. The system developed here successfully connects gesture recognition technology with practical applications for assistive automation.

The system will gain future improvements through several gestures which will expand its functionality to process full sign language sentences and advanced accuracy improvements through transformers and multi-hand and multi-user recognition capabilities. System development should start by programming fundamental functions which enable users to operate the system through its primary functions. The system requires development work to maintain its function when operating in conditions with dim lighting and intricate background environments. Mobile application and cloud platform integration enable users to access the system from remote locations while gesture recognition technology works with voice systems to develop a complete assistive technology solution. The system can control more smart devices which enables it to operate in diverse smart home situations.

REFERENCES

1. R. Yayla, H. Üçgün, and M. Abbas, "Pose-Based Static Sign Language Recognition with Deep Learning for Turkish, Arabic, and American Sign Languages," Sensors, 2026.
2. S. Raut, R. Suryavanshi, P. Jadhav, and S. Patil, "AI-Powered Sign Language Converter Using Deep Learning," IJERT, 2025.
3. R. Alharbi and E. Lim, "Bridging Communication Gaps Using AI: A Review on Sign Language Recognition Systems," Journal of Artificial Intelligence Research, 2024.
4. A. Ahmed, M. Javed, and S. Khan, "End-to-End Sign Language Recognition Using Vision Transformers," IEEE International Conference on Machine Vision and Applications, 2024.
5. L. Huang, H. Wang, and J. Zhao, "SignBERT: Pre-trained Framework for Sign Language Understanding," ACL, 2024.
6. H. Chen, X. Li, and Z. Zhang, "Transformer-Based Multi-Modal Learning for Sign Language Translation," Pattern Recognition Letters, 2024.
7. D. Singh and R. Tiwari, "SignSpeak: AI-Based Real-Time Sign-to-Speech Conversion System," IJEAT, 2024.
8. A. Ghosh, J. Singh, and M. Sharma, "Real-Time Indian Sign Language Recognition Using Mediapipe and CNN," IEEE Conference on Computer Vision Applications, 2023.
9. S. Jain, D. Yadav, and P. Mehta, "3D CNN-Based Indian Sign Language Recognition for Real-Time Translation," IEEE Access, 2023.
10. R. Sharma and S. Gupta, "Enhancing Accessibility Through AI-Driven Sign Language Translators," Springer Advances in Intelligent Systems and Computing, 2023.
11. P. Kumar, V. Nair, and T. George, "Voice-to-Sign Language and Sign- to-Text Translation System Using Deep Neural Networks," International Conference on Artificial Intelligence and Human-Computer Interaction, 2023.
12. R. Kaur, P. Bansal, and T. Sethi, "Speech-to-Sign: A Bidirectional Communication System for the Hearing Impaired," Elsevier Procedia Computer Science, 2023.
13. S. Roy and D. Sen, "Real-Time Indian Sign Language Recognition Using Mediapipe and Deep CNN," ICICT, 2023.
14. M. Suresh Kumar, K. Jayasundar, and S. Ragavendhiran, "Propagation of Sign Language Commands Using NLP," ICCEBS, 2023.
15. Y. Cao, D. Wu, and X. Lin, "Deep Learning-Based Gesture Recognition for Sign Language Translation," Elsevier Neural Computing and Applications, 2022.
16. K. P. Murphy, "Applications of Convolutional Neural Networks in Human Action Recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022.
17. N. Patel and M. Deshmukh, "Design of a Real-Time AI Assistant for Sign-to-Text Conversion Using OpenCV," IJCSIT, 2022.
18. P. Wang and C. Liu, "Gesture-to-Text Conversion Using Deep Recurrent Networks," IEEE

Transactions on Neural Networks and Learning Systems, 2022.

19. L. Gao, Y. Zhang, and F. Wang, "Sign2Text: Transformer-Based Model for Continuous Sign Language Recognition," ACM Multimedia, 2022.
20. M. Koller, O. Zafrir, and S. Ney, "Continuous Sign Language Recognition: Current Trends and Challenges," Computer Speech & Language, vol. 68, 2021.
21. A. Narhe, P. Tade, and C. Mohite, "Vision-Based Gesture Recognition System for Communication Assistance," IJETR, 2020.