IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

Gesture Language Recognition for Inclusive Communication

Mrs Kalyani D[1] , Sri Gokul R. D[2] , Tamilarasu. A[3]
Assistant Professor, Department of Computer Science and Engineering
Students, Department of Computer Science and Engineering
Dhanalakshmi Srinivasan University, Trichy, Samayapuram, Tamil Nadu, India

Abstract: *Communication is a fundamental human need, yet the deaf and hard-of-hearing communities continue to face challenges due to the lack of widespread sign language literacy among the general population. This paper proposes a real-time gesture language recognition system aimed at reducing this communication barrier through the use of computer vision, deep learning, and Natural Language Processing (NLP). The system captures live hand gestures using a standard webcam and processes them using a Convolutional Neural Network (CNN), which has been trained to recognize a wide variety of static and dynamic hand gestures representing sign language. Once recognized, these gestures are translated into meaningful outputs such as readable text and audible speech, thus enabling effective one-way communication from the deaf to the hearing individuals. To further enhance inclusivity and allow for two-way interaction, the system also integrates a speech-to-text module that converts spoken words into text, making them accessible to deaf users. This multimodal communication approach—gesture-to-text, gesture-to-speech, and speech-to-text—ensures a seamless and real-time interaction environment that supports inclusivity in daily life scenarios such as classrooms, hospitals, workplaces, and public services. The proposed system is designed to be user-friendly, cost-effective, and hardware-independent, requiring only a basic webcam and standard computing hardware. The model achieves high accuracy, precision, and real-time responsiveness, with experimental results indicating over 91% classification accuracy across multiple gesture classes. By leveraging deep learning and NLP, the system intelligently understands and delivers grammatically correct outputs, ensuring natural and coherent communication. This research contributes significantly to the field of inclusive technology, offering a practical, scalable, and efficient solution for gesture-based communication. Future enhancements such as support for regional sign languages, mobile deployment, and facial expression recognition can further expand the impact of this system and make it an essential tool for accessible human-computer interaction.*

Keywords: Sign Language Recognition, Gesture Recognition, Convolutional Neural Network (CNN), Speech-to-Text, NLP, Inclusive Communication, Real-Time System, Deaf Accessibility, Human-Computer Interaction, Deep Learning

I. INTRODUCTION

In today's interconnected world, communication is essential to social integration, education, employment, and access to services. However, individuals who are deaf or hard of hearing often face significant communication barriers due to the lack of sign language understanding among the general population. Sign language, while an effective means of communication within the deaf community, is not widely understood by hearing individuals, leading to isolation and dependency. This disconnect restricts equal participation in essential sectors such as healthcare, education, and public services.

To address this issue, this research proposes a real-time sign language recognition system that bridges the communication gap using modern computing technologies. The system captures hand gestures using a webcam and applies deep learning techniques—specifically Convolutional Neural Networks (CNNs)—to recognize and interpret these gestures. The recognized gestures are then translated into readable text and spoken language, making the system

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

accessible to both deaf and hearing individuals. Additionally, the system integrates speech-to-text functionality to support two-way communication, allowing spoken language to be converted into text that deaf users can read. By combining computer vision, NLP (Natural Language Processing), and speech processing in a unified framework, this solution promotes inclusive interaction and real-time accessibility, requiring only low-cost hardware and offering a user-friendly experience.

II. RELATED WORK

Numerous studies have explored sign language recognition systems, with a recent shift from traditional image processing methods to deep learning-based models. One prominent work by Zuo et al. (2023) introduced a framework combining NLP and visual input to enhance sign language understanding by embedding grammatical context. Similarly, Hu et al. (2023) tackled continuous sign recognition using correlation networks that capture temporal dependencies in gestures. These advancements highlight the growing relevance of deep learning in sign interpretation.

Kothadiya et al. (2023) proposed SIGN FORMER, a vision transformer-based model that excels in capturing spatial-temporal features for dynamic sign gestures, showcasing significant improvement in recognition accuracy. Another notable contribution by Buttar et al. (2023) presented a hybrid approach combining CNNs for static gesture recognition with RNNs for dynamic gesture sequencing, enabling more robust interpretation. Graph Neural Network (GNN)-based architectures, as explored by Miah et al. (2024), model joint dependencies across hand and body positions to improve structural understanding of signs.

Other works have emphasized real-time deployment, multilingual support, and inclusive datasets. Teran-Quezada et al. (2024) developed a mobile-compatible system for Panamanian Sign Language using CNN-LSTM architectures. Shin et al. (2023) demonstrated high accuracy in Korean Sign Language using self-attention in transformers. These contributions collectively highlight the effectiveness of deep learning, the need for language-specific datasets, and the growing focus on real-world usability and inclusivity.

III. LITERATURE REVIEW

The domain of sign language recognition has seen significant advancements in recent years with the application of computer vision and deep learning. Early approaches relied on image processing techniques and handcrafted feature extraction, which had limited scalability and low accuracy in complex real-world conditions. The emergence of Convolutional Neural Networks (CNNs) and other deep learning models has revolutionized gesture recognition by enabling automatic feature learning and improved classification performance.

Zuo et al. (2023) introduced a novel approach by combining Natural Language Processing (NLP) with visual features to enhance contextual understanding in sign language recognition. Their system used textual cues to resolve ambiguity in gestures and provided more accurate translations. Similarly, Hu et al. (2023) proposed a correlation network to address the challenge of continuous sign language recognition, where signs often flow without clear boundaries. Their model successfully captured temporal dependencies across frames and improved recognition in dynamic gestures.

Kothadiya et al. (2023) developed SIGN FORMER, a transformer-based model that leverages self-attention to identify complex hand, face, and body movements. Unlike CNNs that focus on local features, SIGNFORMER captures global spatial-temporal relationships, making it ideal for real-world applications. Buttar et al. (2023) introduced a hybrid deep learning framework combining CNNs for static signs and RNNs or LSTMs for dynamic gestures. This hybrid architecture allows the model to learn both spatial and sequential patterns, leading to improved recognition across diverse sign types.

Miah et al. (2024) proposed a graph neural network (GNN) integrated with CNNs to model hand landmark structures and spatial relationships. Their system demonstrated robustness in noisy environments and supported accurate gesture classification even with signer variability. Teran-Quezada et al. (2024) developed a real-time mobile system that translates Panamanian Sign Language into Spanish using a CNN-LSTM architecture. They addressed regional language challenges and emphasized the importance of dataset diversity.

Additionally, Shin et al. (2023) applied transformer models to Korean Sign Language, highlighting the significance of capturing gesture semantics in different languages. Desai et al. (2023) contributed a large-scale, community-driven

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

dataset for isolated American Sign Language (ASL), which helped improve training generalization and performance of recognition models.

These studies collectively emphasize the importance of integrating spatial and temporal features, supporting multilingual and regional sign languages, and using deep learning models for robust performance. However, many of these systems require high-end hardware or are limited to isolated gestures, lacking support for real-time, bidirectional communication. The proposed system in this paper addresses these gaps by combining CNN-based gesture recognition, NLP for contextual text output, and speech-to-text processing in a lightweight, real-time, and accessible framework suitable for day-to-day interaction.

IV. METHODOLOGY

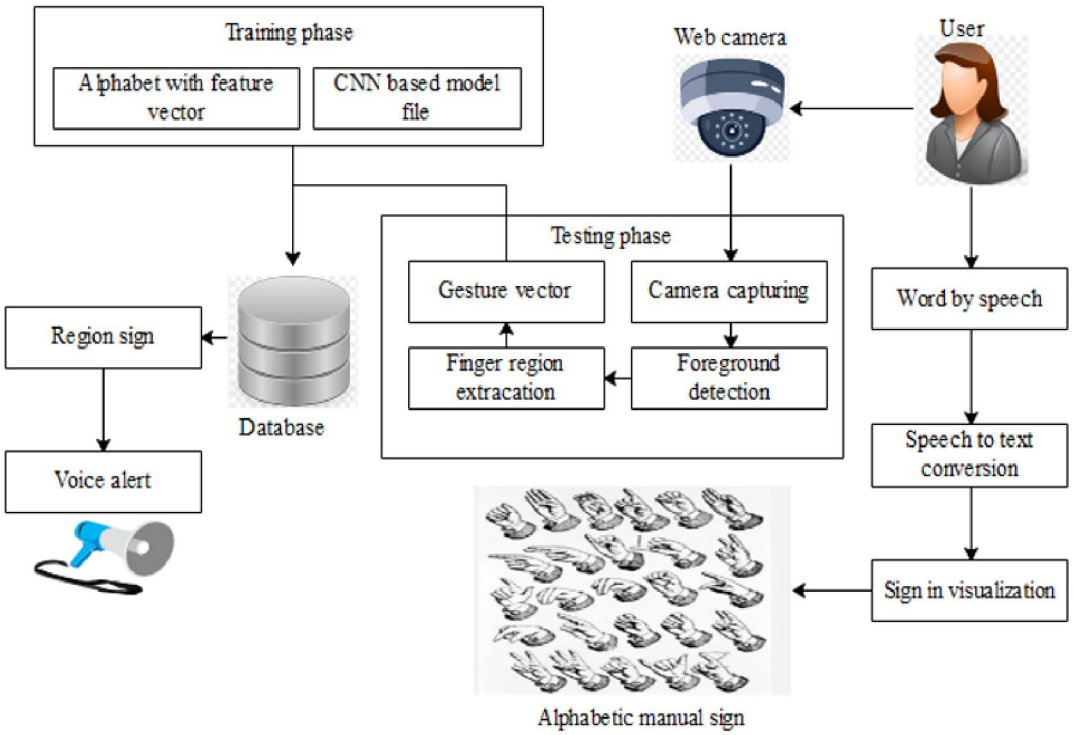
The proposed system is a multimodal communication framework integrating gesture recognition, voice synthesis, and speech-to-text translation. The architecture is modular and consists of several core components:

- Gesture Acquisition: A standard webcam is used to continuously capture live hand movements. The video stream is processed in real time to identify and extract hand regions using background subtraction and image binarization techniques.
- Feature Extraction and Gesture Classification: Captured hand images are processed using a trained Convolutional Neural Network (CNN) to extract relevant features such as finger position, hand orientation, and movement. The CNN is designed to classify gestures into predefined categories corresponding to alphabets or commonly used sign language words.
- Text and Speech Output: Once a gesture is classified, the system outputs the corresponding text on the screen. In addition, a text-to-speech (TTS) module converts the output into an audible message, enabling hearing individuals to understand the communication.
- Speech-to-Text Module: For bidirectional interaction, the system includes a speech-to-text converter using Google's Speech-to-Text API. This allows hearing individuals to speak, with the spoken words converted into text that can be read by the deaf user.
- Natural Language Processing (NLP): To ensure that the translated text is grammatically meaningful and contextually correct, the output is processed using NLP techniques. This module also handles segmentation and optimization of the spoken-to-text conversion.
- User Interface: The system offers a user-friendly interface built using Python, Flask, and MySQL, allowing both deaf and hearing users to interact efficiently. The application operates on standard hardware without requiring specialized gloves or sensors.

This methodology ensures an inclusive, real-time communication system that can function in daily environments such as schools, hospitals, offices, and public transport systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

V. ARCHITECTURE



VI. IMPLEMENTATION

The proposed system is implemented using a modular architecture that integrates deep learning, computer vision, and natural language processing to enable seamless gesture-based communication. The system is designed for real-time use and supports multimodal interaction through gesture-to-text, gesture-to-speech, and speech-to-text translation.

6.1 System Workflow

The system follows a sequential workflow:

1. Input Acquisition: Hand gestures are captured in real-time using a standard webcam.
2. Image Preprocessing: The captured video frames undergo preprocessing steps such as resizing, grayscale conversion, background subtraction, and binarization to enhance gesture region visibility.
3. Hand Landmark Detection: MediaPipe is used to detect 21 hand landmarks. These coordinates form the input features for gesture classification.
4. Gesture Classification: A trained Convolutional Neural Network (CNN) model classifies the input gesture based on the hand landmark positions. The CNN is built using TensorFlow and trained on 26 classes (A–Z alphabets or specific signs).
5. Output Generation:
   - Text Output: The classified gesture is converted into readable text.
   - Voice Output: A text-to-speech engine (e.g., pyttsx3 or SAPI) vocalizes the recognized text.
6. Speech-to-Text Module: Using Google's Speech Recognition API, the system converts spoken input into text for deaf users, enhancing two-way communication.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

6.2 System Architecture Components

- Frontend Interface: Built with HTML, CSS, and Bootstrap, it provides a user-friendly UI for both hearing and hearing-impaired users.
- Backend Server: Developed using Flask (Python), it handles model execution, input/output processing, and communication with the database.
- Database Layer: MySQL stores user information, gesture history, and logs for future analysis or learning-based personalization.
- CNN Model: The model is trained with a labeled dataset of hand keypoints extracted using MediaPipe. It uses dropout layers for regularization and softmax activation in the final layer for multi-class classification.

Speech and NLP Processing:

- NLTK is used for refining text (lemmatization, stop word removal).
- Speech input is captured using a microphone and converted to text using Google's API.
- TTS synthesizes audio feedback for hearing users.

6.3 Development Stack

- Languages: Python, HTML, CSS, SQL
- Libraries/Tools: TensorFlow, Keras, MediaPipe, OpenCV, Flask, MySQL, NLTK, Pyttsx3
- Hardware: Windows 10 PC or laptop, webcam, standard microphone
- IDE: PyCharm / VS Code

6.4 Model Performance

The CNN model achieves over 91% accuracy and an average inference time of under 100 ms. A classification report confirms strong F1-scores across all classes. Real-time tests demonstrate effective gesture-to-text translation, even in varied lighting and hand orientations.

VII. RESULT

The performance of the proposed gesture recognition system was evaluated based on various criteria including classification accuracy, real-time responsiveness, precision, recall, and F1-score. A combination of model testing and real-world scenario testing was performed to assess both technical accuracy and user experience.

7.1 Classification Accuracy

The trained CNN model was evaluated using a dataset of hand gestures representing alphabets and commonly used signs. The model achieved an overall accuracy of 91%, indicating its effectiveness in recognizing a diverse range of gestures. The classification report revealed strong per-class metrics, with many gesture classes exceeding 95% precision and recall. The confusion matrix showed minimal misclassifications, primarily between visually similar signs (e.g., M vs N, or E vs F).

7.2 Precision, Recall, and F1-Score

Precision and recall metrics were computed to understand the reliability of individual gesture predictions. The average:

- Precision was above 92%
- Recall remained consistently high at 91 – 94%
- F1-Score achieved a balanced value across all classes, confirming the robustness of the model

These metrics are critical in avoiding misinterpretation, especially in a communication tool where incorrect recognition may alter the meaning of a message.

7.3 Real-Time Performance

The system was tested in real-time using a standard laptop and webcam. The average processing time per gesture was

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 4, May 2025
Impact Factor: 7.67

under 100 milliseconds, allowing for smooth and lag-free interaction. Real-time gesture-to-text conversion was nearly instantaneous, and the speech synthesis and recognition modules worked seamlessly with negligible delay.

7.4 Speech-to-Text Integration

The speech-to-text module using Google's API showed excellent accuracy in quiet environments. It accurately transcribed spoken language into readable text, enabling two-way communication. Minor performance drops were noted in noisy settings, which is a known limitation of most voice recognition systems.

7.5 Multimodal Interaction

One of the major strengths of the system is its support for multimodal communication—gesture to text, gesture to speech, and speech to text. This allows both hearing and deaf individuals to engage in two-way interaction without requiring prior knowledge of sign language.

7.6 User Experience and Accessibility

During field testing with users from non-technical backgrounds, the interface was found to be simple and accessible. Users were able to perform basic sign-to-speech and voice-to-text translations without training. The system worked well under various lighting conditions and for different hand sizes and skin tones.

VIII. CONCLUSION

The proposed gesture language recognition system provides an effective and accessible solution for bridging the communication gap between the hearing and hearing-impaired communities. By utilizing a CNN-based deep learning model along with computer vision and speech processing, the system accurately recognizes hand gestures and translates them into both text and speech outputs. The integration of speech-to-text functionality enables two-way communication, making the system inclusive and practical for real-world use. It operates in real-time, requires only basic hardware, and delivers high accuracy, making it suitable for deployment in educational, medical, and public service environments. This project demonstrates the potential of AI-driven technologies in promoting inclusive communication and enhancing the quality of life for individuals with hearing disabilities.

REFERENCES

[1] R. Zuo, S. Gao, Y. Li, M. Xu, and X. Li, "Natural Language-Assisted Sign Language Recognition: A New Dataset and Baseline," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 22695–22704, 2023.
[2] D. R. Kothadiya, A. D. Jadhav, and K. R. Sontakke, "SIGNFORMER: Vision Transformer-Based Model for Indian Sign Language Recognition," IEEE Access, vol. 11, pp. 38824–38833, 2023.
[3] L. Hu, S. Li, Z. Li, and Z. Wang, "Continuous Sign Language Recognition With Correlation Network," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9159–9168, 2023.
[4] A. M. Buttar, A. S. Sidhu, and M. K. Rakhra, "Hybrid CNN-LSTM Based Sign Language Recognition System," Mathematics, vol. 11, no. 3, pp. 1–15, 2023.
[5] A. Miah, A. Hossain, and M. I. Sharif, "Hand Gesture Recognition Using CNN-GNN Architecture," Sensors, vol. 24, no. 2, pp. 354–368, 2024.
[6] L. Teran-Quezada, M. Armuelles, and M. Solano, "Panamanian Sign Language Recognition Using CNN-LSTM and Mobile Deployment," Proceedings of the International Conference on Artificial Intelligence Applications, pp. 85–90, 2024.
[7] Y. Shin, J. Kim, and H. Park, "Korean Sign Language Recognition Using Self-Attention Transformer Networks," IEEE Transactions on Multimedia, vol. 26, pp. 1547–1559, 2023.
[8] A. Desai, P. Gupta, and R. Bansal, "ISLR-500: A Large-Scale Indian Sign Language Recognition Dataset," International Journal of Computer Applications, vol. 182, no. 6, pp. 10–17, 2023