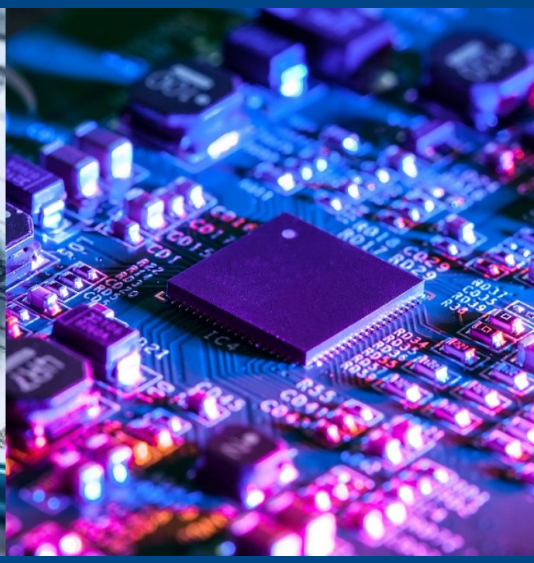
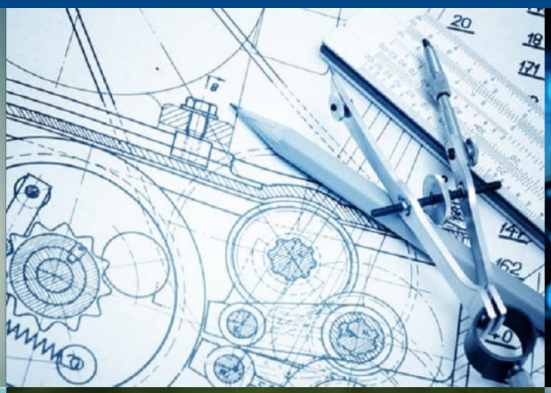
ISSN: 2582-7219

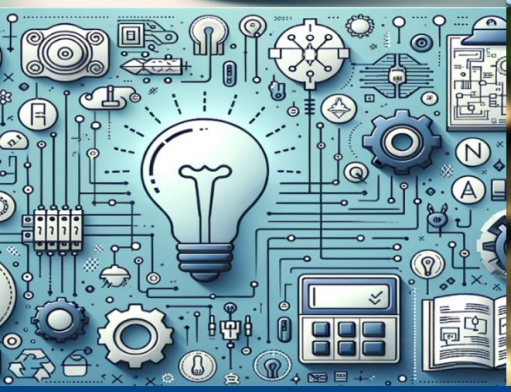




International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206 **Volume 9, Issue 4, April 2026**

© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904300

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Conversion of Sign Language into Text and Speech using Deep Learning

Abdul Khadar[1], Akif Rayan Sherif[2], Imran Farhat[3], Nayeem[4], Kiruthiga R[5]

Student, Department of Artificial Intelligence and Data Science, Aalim Muhammed Salegh College of Engineering, Avadi, Chennai, Tamil Nadu, India[1]

Student, Department of Artificial Intelligence and Data Science, Aalim Muhammed Salegh College of Engineering, Avadi, Chennai, Tamil Nadu, India[2]

Student, Department of Artificial Intelligence and Data Science, Aalim Muhammed Salegh College of Engineering, Avadi, Chennai, Tamil Nadu, India[3]

Student, Department of Artificial Intelligence and Data Science, Aalim Muhammed Salegh College of Engineering, Avadi, Chennai, Tamil Nadu, India[4]

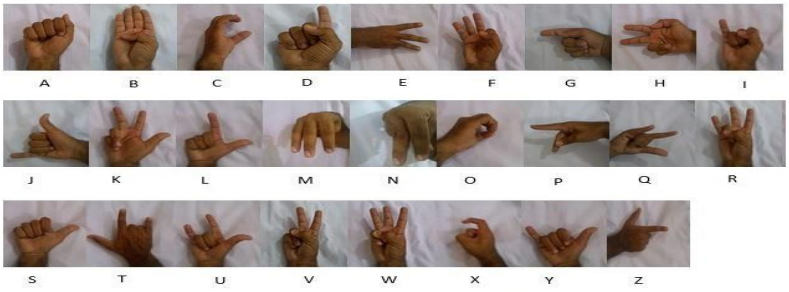
Assistant Professor, Department of Artificial Intelligence and Data Science, Aalim Muhammed Salegh College of Engineering, Avadi, Chennai, Tamil Nadu, India[5]

ABSTRAT: Sign language serves as the primary means of communication for deaf and speech-impaired individuals, yet it remains inaccessible to a majority of the population. This communication gap necessitates the development of automated systems capable of interpreting hand gestures into understandable formats. This paper presents a vision-based sign language recognition system that converts hand gestures into text and speech using image processing and deep learning techniques. The proposed system employs pre-processing methods such as segmentation and morphological filtering, followed by feature extraction and classification using Linear Discriminant Analysis (LDA) and Convolutional Neural Networks (CNN). The system is designed to operate in real time using a webcam and demonstrates improved accuracy and efficiency compared to traditional approaches. The results indicate that computer interaction for assistive communication.

KEYWORDS: Sign Language Recognition, Computer Vision, Convolutional Neural Networks (CNN), Gesture Recognition, Real-Time Systems, Assistive Communication, Human-Computer Interaction, Image Processing

I. INTRODUCTION

Sign language is a visual means of communication that utilizes hand gestures, facial expressions, and body movements. It is widely used by deaf and speech-impaired individuals; however, it is not commonly understood by the general population. This lack of mutual understanding creates a significant communication barrier, limiting interaction and social inclusion. Bridging this gap requires an intelligent system capable of translating sign language into text or speech in real time.With the advancement of computer vision and machine learning, automated gesture recognition has become a feasible solution. Traditional systems relied on hardware-based approaches such as sensor gloves, which, although accurate, are inconvenient and expensive. Vision-based approaches using cameras and image processing techniques



© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904300

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

offer a more practical and scalable solution.Recent developments in deep learning, particularly Convolutional Neural Networks (CNN), have significantly improved the accuracy of image classification tasks. These models can automatically extract features from images, eliminating the need for manual feature engineering. However, challenges remain in achieving high accuracy under varying lighting conditions, background noise, and gesture variations.

This paper proposes a hybrid approach combining classical image processing techniques with deep learning methods to develop an efficient sign language recognition system. The system is designed to capture gestures, process them, extract meaningful features, and classify them into corresponding text and speech outputs, thereby enabling seamless communication between deaf individuals and others.

II. BACKGROUND AND RELATED WORK

Sign Language Recognition Systems

Sign language recognition has been an active area of research, with various approaches proposed over the years. Early systems primarily relied on glove-based techniques, where sensors attached to the hand captured gesture information. While these systems provided high accuracy, they required specialized hardware, making them impractical for everyday use.Vision-based approaches emerged as a more accessible alternative, utilizing cameras to capture gestures and applying image processing techniques for recognition. These methods eliminate the need for wearable devices but introduce challenges related to background noise, lighting variations, and segmentation accuracy.

Machine Learning Approaches

Traditional machine learning methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Euclidean distance-based classifiers have been widely used for gesture recognition. These methods rely on manual feature extraction and are effective for simple classification tasks.With the advancement of deep learning, Convolutional Neural Networks (CNN) have become the preferred approach forimage-based recognition tasks. CNN models automatically learn hierarchical features from data, leading to improved accuracy and robustness. They have been successfully applied in various domains, including facial recognition, object detection, and gesture recognition.

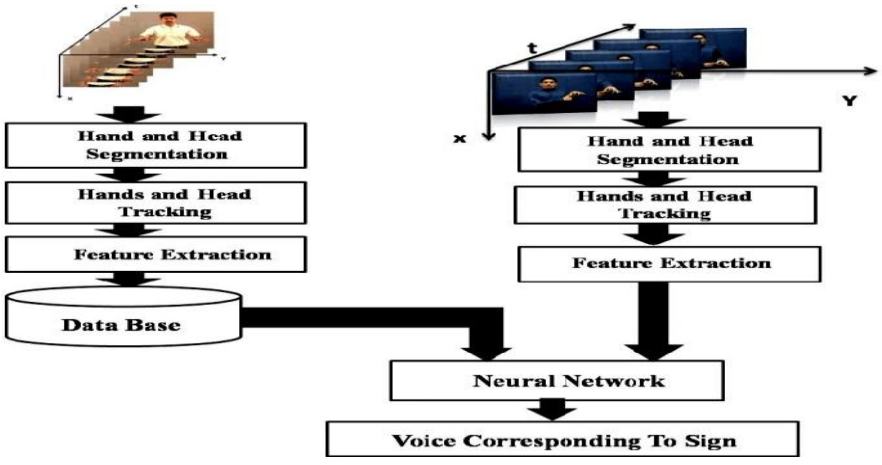
Image Processing Techniques

Image processing plays a crucial role in sign language recognition systems. Techniques such as grayscale conversion, segmentation, and morphological filtering are used to enhance image quality and isolate the region of interest. Otsu's thresholding method is commonly used for segmentation, as it effectively separates the foreground (hand) from the background.

III. PROPOSED SYSTEM

System Overview

The proposed system is designed as a vision-based framework that converts hand gestures into text and speech. It consists of multiple interconnected stages, including data acquisition, pre-processing, feature extraction, and classification. These stages work together to ensure accurate and real-time gesture recognition.



© 2026 IJMRSET | Volume 9, Issue 4, April 2026 | DOI:10.15680/IJMRSET.2026.0904300

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The system captures hand gestures using a webcam and processes the input images to extract meaningful features.

These features are then classified using machine learning and deep learning algorithms to determine the corresponding sign. Finally, the recognized gesture is converted into text and speech output.

Data Acquisition

The data acquisition stage involves capturing hand gesture images using a webcam. Multiple samples of each gesture are collected under different conditions to improve the robustness of the system. These images are stored in a dataset and used for training and testing the model. Proper lighting and positioning are maintained to ensure clarity and consistency in the captured images.

Pre-processing

Pre-processing is performed to prepare the captured images for further analysis. The images are first converted into grayscale to reduce computational complexity. Segmentation is then applied using Otsu's thresholding method to separate the hand region from the background.

Morphological filtering techniques such as erosion and dilation are used to remove noise and enhance the shape of the hand. This results in a clean and well-defined image that can be used for accurate feature extraction.

Feature Extraction

- Feature extraction is a critical step in the recognition process. In this stage, important characteristics of the hand gesture are extracted from the pre-processed images.
- Techniques such as Eigenvalue and Eigenvector computation are used to represent the gesture in a reduced-dimensional space.
- This transformation helps in capturing the essential features of the gesture while minimizing redundant information, thereby improving classification efficiency.

Classification

The classification stage involves identifying the gesture based on the extracted features. Linear Discriminant Analysis (LDA) is used to perform dimensionality reduction and enhance class separability. Additionally, Convolutional Neural Networks (CNN) are employed to improve recognition accuracy by automatically learning complex patterns from the data.The trained model predicts the corresponding gesture label, which is then mapped to its respective text and speech output.

IV. IMPLEMENTATION AND EVALUATION

Implementation Details

The system is implemented using Python, leveraging libraries such as OpenCV for image processing, TensorFlow and Keras for deep learning, and NumPy for numerical computations. The webcam is used for real-time image capture, and the processing pipeline is executed sequentially to generate output.The model is trained using a dataset of hand gesture images, with a portion reserved for testing. The training process involves optimizing the model parameters to minimize classification error.

Experimental Evaluation

The system is evaluated based on metrics such as accuracy, precision, recall, and F1-score. The results demonstrate that the proposed system achieves high accuracy in recognizing hand gestures under controlled conditions. The integration of CNN significantly improves performance compared to traditional methods. The system also shows good real-time performance, making it suitable for practical applications.

V. CONCLUSION

This paper presented a vision-based sign language recognition system that converts hand gestures into text and speech using image processing and deep learning techniques. By combining classical methods such as LDA with modern approaches like CNN, the system achieves improved accuracy and efficiency.The proposed system successfully bridges the communication gap between deaf individuals and the general population, enabling more inclusive interaction.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Future work will focus on extending the system to support dynamic gestures, sentence-level recognition, and deployment on mobile platforms for wider accessibility.

VI. ACKNOWLEDGEMENT:

The authors would like to thank the faculty and the Department of Artificial Intelligence and Data Science at Aalim Muhammed Salegh College of Engineering for their guidance and support throughout this project.

REFERENCES

[1] Mahesh Kumar N B, Conversion of Sign Language into Text, International Journal of Applied Engineering Research (2018).
[2] Jebakani C., Conversion of Sign Language into Speech using CNN, Sathyabama Institute of Science and Technology (2022).

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com