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 6, Issue 3, November 2026
Impact Factor: 8.2

Vision-Based Real-Time Gesture to Speech & Speech to Text Translation for Sign Language

Miss. Pooja Jori[1], Miss. Khushi Hiwarkar[2], Miss. Sanika Gadge[3], Prof. Madhavi Kolse[4]
Students, Department of Artificial Intelligence and Data Science[123]
Professor, Department of Artificial Intelligence and Data Science[4]
JCEI's Jaihind College of Engineering, Kuran, Maharashtra, India

Abstract: *The proposed system aims to reduce communication barriers for individuals with hearing or speech impairments using a real-time, vision-based gesture-to-speech and speech-to-text translation system. It captures hand gestures through a webcam, processes images using OpenCV, and recognizes gestures using a Convolutional Neural Network (CNN) to generate text and speech through pyttsx3. Additionally, spoken input is converted into text using the Speech-Recognition library. The system integrates both functionalities into a single application to enable accessible and effective communication between sign language users and non-sign language users. By providing two-way communication support, the proposed approach helps users interact more conveniently in everyday situations. The use of real-time image processing enables continuous gesture detection and recognition. Furthermore, the system aims to provide a simple, user-friendly, and cost-effective solution without requiring specialized hardware. This integrated approach has the potential to improve accessibility in educational institutions, workplaces, healthcare environments, and other public spaces.*

Keywords: Convolutional Neural Network

I. INTRODUCTION

Communication is an essential part of human life, as it allows individuals to express their thoughts, exchange information, and interact with others. However, people with hearing and speech impairments often experience communication difficulties when interacting with individuals who are unfamiliar with sign language. Sign language provides an effective means of expression, but the lack of awareness and understanding among the general population creates communication barriers in everyday situations. With the development of Artificial Intelligence (AI) and computer vision, technology can be used to make communication more accessible and convenient. The proposed system introduces a vision-based approach that recognizes hand gestures and converts them into meaningful text and speech. It uses a webcam to capture hand movements, while OpenCV performs image acquisition and preprocessing. A Convolutional Neural Network (CNN) is employed to identify and classify the captured gestures based on the trained dataset. The recognized gestures are displayed as text and converted into audible speech using the pyttsx3 library. In addition, the system provides a speech-to-text feature through the Speech-Recognition library, allowing spoken words to be converted into readable text. By combining These two Functions, the application supports communication in Both Direction .The main Goal is to develop an accissible , user friendly, and affordable communication solution they can help reduce the communication gap.

II. OBJECTIVE

1. To develop a real-time vision-based system that recognizes sign language gestures and converts them into meaningful speech.
2. To implement accurate hand gesture detection and classification using computer vision and deep learning techniques.
3. To enable speech-to-text conversion, allowing spoken communication to be displayed as readable text.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-40809

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 6, Issue 3, November 2026
Impact Factor: 8.2

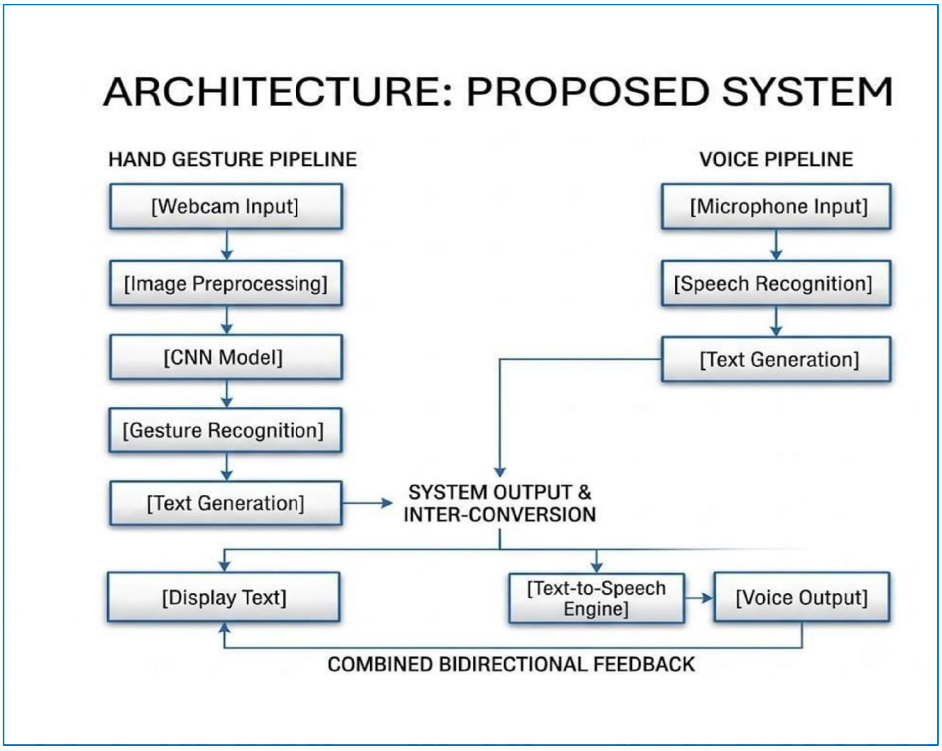
4. To create an accessible, user-friendly communication platform that reduces interaction barriers between hearing-impaired and non-sign-language users.

III. LITERATURE SURVEY

1. Sandberg (1997): Sandberg proposed a gesture recognition technique using neural networks, demonstrating the potential of artificial intelligence for recognizing hand gestures.
2. Shashidhar et al. (2022): The researchers developed an Indian Sign Language recognition system using a 2-D Convolutional Neural Network (CNN) with a graphical user interface for real-time gesture recognition.
3. Kumar et al. (2023): The authors introduced a lightweight deep-learning architecture for sign language classification, focusing on efficient and faster gesture recognition.
4. Gogoi et al. (2025): The researchers presented a vision-based real-time gesture-to-speech translation system using an optimized CNN and Pyttsx3, providing a foundation for converting recognized gestures into text and speech.

IV. METHODOLOGY

1. Image Acquisition: Capturing hand gestures using a webcam in real time.
2. Image Processing: Using OpenCV to detect and process hand movements.
3. Gesture Recognition: Applying a trained CNN model to identify sign language gestures.
4. Gesture-to-Speech: Converting recognized gestures into text and audible speech.
5. Speech-to-Text: Converting spoken words into readable text using speech recognition.
6. Output: Displaying text and generating voice output for effective two-way communication.



V. SYSTEM ARCHITECTURE

Hardware Requirement :

1. Laptop/Computer: Runs the application and performs processing tasks.

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 6, Issue 3, November 2026 Impact Factor: 8.2

2. Webcam: Captures hand gestures in real time.
3. Microphone: Records spoken input for speech-to-text conversion.
4. Speakers: Produces audio output for recognized gestures.
5. Processor (Intel i3 or above): Handles image processing and application execution.
6. RAM (Minimum 4 GB): Supports smooth execution of the application.

Software Requirement :

1. Python: Main programming language used to develop the system.
2. OpenCV: Captures and processes real-time hand gesture images.
3. TensorFlow/Keras: Used to build and train the CNN model.
4. NumPy: Performs numerical operations and array processing.
5. SpeechRecognition: Converts spoken input into text.
6. pyttsx3: Converts recognized gesture text into speech.
7. VS Code / Jupyter Notebook: Used for coding and project development.
8. Operating System (Windows): Provides the platform to run the application.

VI. RESULTS

The proposed system successfully integrates sign language recognition and speech-to-text conversion into a single application. The webcam captures hand gestures, which are processed using OpenCV and classified through a CNN model. The recognized gestures are displayed as text and converted into speech using pyttsx3. Similarly, spoken input is converted into readable text using the SpeechRecognition library. The system provides an effective approach to improving communication between sign language users and the general public.

VII. BENEFITS OF SOCIETY

The proposed system provides several benefits to society by promoting effective communication between individuals with hearing or speech impairments and the general public. It helps reduce communication barriers through real-time gesture recognition and speech conversion. The system supports social inclusion, improves educational accessibility, and encourages independent communication in daily activities. Its cost-effective and user-friendly approach can make communication more convenient in educational institutions, workplaces, healthcare centres, and public places.

VIII. CONCLUSION

The proposed system presents an effective approach to reducing communication barriers for individuals with hearing and speech impairments. By integrating Sign Language Recognition, Gesture-to-Speech Translation, and Speech-to-Text Conversion, it enables two-way communication through a single application. The combination of OpenCV, CNN, and speech processing technologies makes the system useful for real-time interaction. This approach promotes accessibility, independence, and social inclusion, contributing towards a more inclusive and technology-driven society.

IX. FUTURE SCOPE

The proposed system can be further improved by supporting multiple sign languages and a larger vocabulary of gestures. Advanced deep learning techniques can be integrated to improve recognition accuracy and processing speed. Future development may include mobile application support, facial expression recognition, and continuous sentence formation. The system can also be enhanced to work in different lighting conditions and noisy environments, making it more reliable, accessible, and user-friendly for real-world applications.

X. ACKNOWLEDGEMENT

We would like to express our sincere gratitude to our project guide and faculty members for their valuable guidance, continuous support, and encouragement throughout this project. We are thankful to our institution for providing the

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 6, Issue 3, November 2026
Impact Factor: 8.2

necessary resources and facilities. We also extend our thanks to everyone who directly or indirectly contributed to the successful completion of this research work.

REFERENCES

[1] N. C. Camgoz, S. Hadfield, O. Koller, and R. Bowden, "SubUNets: End-to-End Hand Shape and Continuous Sign Language Recognition," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
[2] O. Koller, H. Ney, and R. Bowden, "Deep Hand: How to Train a CNN on 1 Million Hand Images When Your Data Is Continuous and Weakly Labelled," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
[3] N. Buckley, L. Sherrett, and E. L. Secco, "A CNN Sign Language Recognition System with Single & Double-Handed Gestures," 2021.
[4] S. A. M. A. S. Senanayaka et al., "Continuous American Sign Language Recognition Using Computer Vision and Deep Learning Technologies," 2022.
[5] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000.