IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁙ Peer-reviewed & Refereed journal ⁙ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151138

Gesture Recognition for Voice Synthesis

BHARGAV K[1], SANDARSH GOWDA M M[2]

Department of MCA, BIT, K.R. Road, V.V. Pura, Bangalore, India[1]

Assistant Professor, Department of MCA, BIT, K.R. Road, V.V. Pura, Bangalore, India[1]

Abstract: Gesture recognition for voice synthesis is an emerging assistive technology that enables communication through hand gestures by converting them into synthesized speech. This system is especially beneficial for individuals with speech or hearing impairments, providing them an alternative medium for interaction. With advancements in computer vision and deep learning, gesture-based interfaces have become more accurate and efficient.

This paper presents a **Gesture Recognition for Voice Synthesis System** that uses computer vision and deep learning techniques to recognize predefined hand gestures in real time and convert them into corresponding voice outputs. A Convolutional Neural Network (CNN) is employed for gesture classification after preprocessing steps such as image resizing, background normalization, and feature extraction. Once a gesture is recognized, a text-to-speech module generates an appropriate voice output.

The system is implemented as a real-time application using a camera interface, allowing users to perform gestures naturally. Experimental evaluation shows high recognition accuracy and low response latency, demonstrating the effectiveness of the proposed system for assistive communication and human–computer interaction applications.

Keywords: Gesture Recognition, Voice Synthesis, Computer Vision, Deep Learning, CNN, Assistive Technology

I. INTRODUCTION

Human–computer interaction has evolved significantly with the introduction of gesture-based systems that allow users to communicate naturally without physical contact. Gesture recognition combined with voice synthesis offers a powerful solution for enabling communication, especially for individuals who are unable to speak.

Traditional communication aids rely on manual input or predefined hardware devices, which can be slow and inconvenient. Vision-based gesture recognition systems overcome these limitations by using cameras and intelligent algorithms to interpret gestures in real time. By mapping recognized gestures to synthesized speech, the system creates an intuitive and efficient communication platform.

This project presents a **Gesture Recognition for Voice Synthesis System** that detects hand gestures using a camera, classifies them using deep learning models, and converts them into voice outputs using text-to-speech technology. The system focuses on accuracy, real-time performance, and ease of use.

1.1 Project Description

The project aims to design and implement a gesture-based voice synthesis system that recognizes predefined hand gestures and converts them into spoken words or sentences. The system captures gesture images through a camera, preprocesses them, and feeds them into a CNN-based classification model.

Once a gesture is successfully recognized, the corresponding text label is passed to a text-to-speech engine, which produces an audible voice output. The system supports real-time interaction and can be extended to include additional gestures or languages.

1.2 Motivation

Individuals with speech impairments often face difficulties in communicating their thoughts effectively. Existing solutions may require typing, specialized hardware, or prior training, which can limit accessibility.

DOI: 10.17148/IJARCCE.2026.151138

This project is motivated by the need to develop a **low-cost, vision-based, and user-friendly communication system** that allows users to express themselves using natural hand gestures. By integrating gesture recognition with voice synthesis, the system reduces dependency on traditional input methods and improves communication efficiency.

II. RELATED WORK

Paper [1] explores vision-based hand gesture recognition using traditional image processing techniques and highlights challenges related to lighting and background noise.

Paper [2] presents machine learning approaches such as SVM and KNN for gesture classification, emphasizing feature extraction techniques.

Paper [3] discusses CNN-based deep learning models for hand gesture recognition and reports improved accuracy over traditional methods.

Paper [4] focuses on real-time gesture recognition systems for assistive communication and evaluates system latency and usability.

Paper [5] reviews gesture-to-speech systems and concludes that deep learning-based approaches provide better scalability and robustness.

III. METHODOLOGY

A. System Environment

The proposed system operates in a real-time environment using a standard camera for gesture capture. Users perform hand gestures in front of the camera, and the system processes the input frames to recognize gestures and generate voice output.

The system is designed to handle varying lighting conditions and background environments while maintaining stable performance. This setup simulates real-world usage scenarios for assistive communication applications

B. System Architecture

Client-Side Processing:
The camera continuously captures video frames containing hand gestures. The frames are converted into images and passed through preprocessing steps such as resizing, grayscale conversion, and background normalization.

Server-Side Execution: The processed images are fed into a trained CNN model for gesture classification. The predicted gesture label is mapped to predefined text. A text-to-speech module converts the text into synthesized speech, which is played back to the user in real time.

C. Gesture Recognition and Voice Synthesis Mechanism

The system dynamically recognizes gestures based on trained data samples. The CNN model extracts spatial features from hand images and classifies them accurately. The recognized gesture is translated into text and converted into speech using a voice synthesis engine. This modular design allows easy integration of new gestures and voice outputs.

D. Implementation Flow

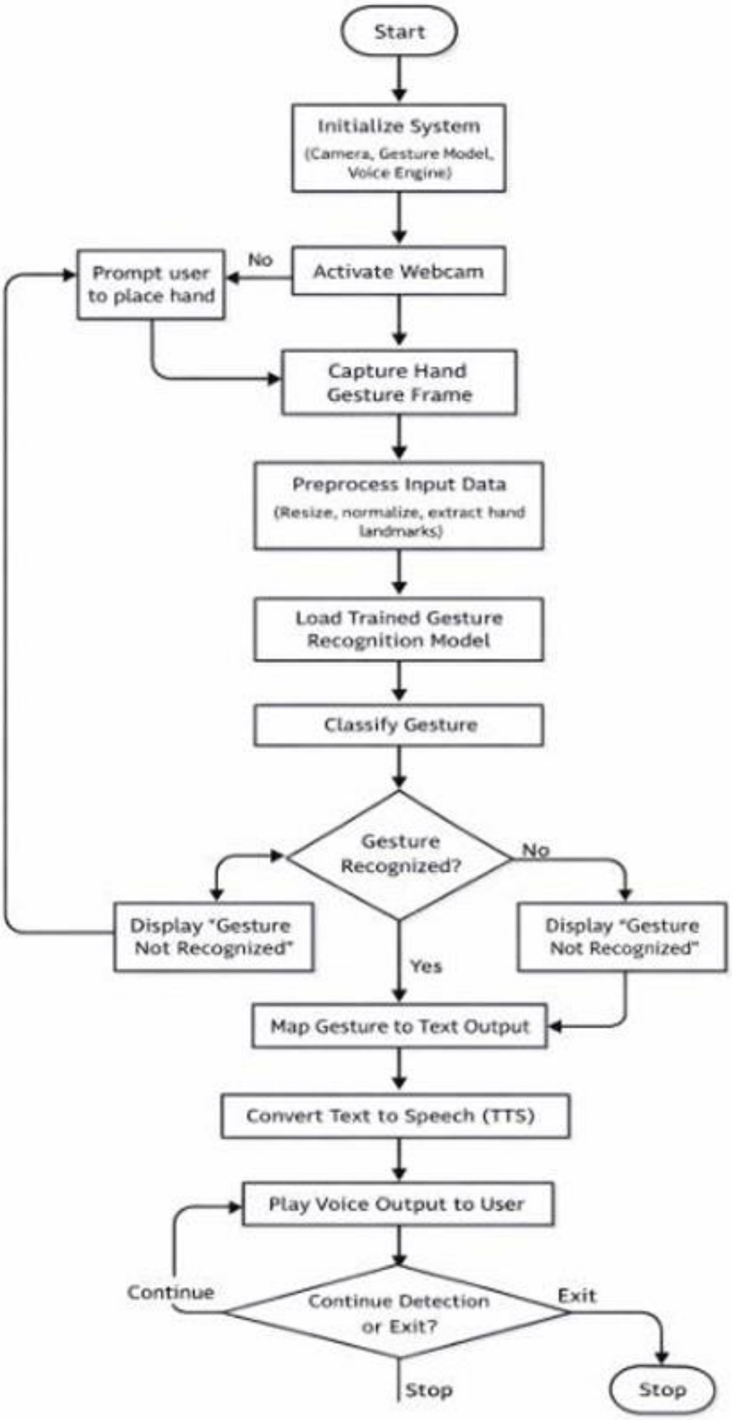
1. The camera captures real-time gesture images.
2. Image preprocessing is applied for noise reduction and normalization.
3. The CNN model classifies the gesture.
4. The recognized gesture is mapped to text.
5. Text-to-speech conversion generates voice output.
6. The synthesized voice is played to the user.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.151138


Start
Initialize System (Camera, Gesture Model, Voice Engine)
Activate Webcam
No
Prompt user to place hand
Capture Hand Gesture Frame
Preprocess Input Data (Resize, normalize, extract hand landmarks)
Load Trained Gesture Recognition Model
Classify Gesture
Gesture Recognized?
No
Display "Gesture Not Recognized"
Display "Gesture Not Recognized"
Yes
Map Gesture to Text Output
Convert Text to Speech (TTS)
Play Voice Output to User
Continue
Continue Detection or Exit?
Exit
Stop
Stop

Hardware:
The system requires a minimum of 8 GB RAM and a standard processor for smooth operation. A GPU is optional and mainly used during the model training phase to improve training speed.

Software:
The Gesture Recognition for Voice Synthesis system is developed using Python as the primary programming language. OpenCV is used for real-time gesture detection and image preprocessing, while TensorFlow and Keras are employed to build and train the CNN model for gesture classification. Text-to-speech libraries such as pyttsx3 or gTTS are integrated to convert recognized gestures into synthesized voice output.

DOI: 10.17148/IJARCCE.2026.151138

IV. SIMULATION AND EVALUATION FRAMEWORK

The system is evaluated based on gesture recognition accuracy, response time, and reliability. Multiple test cases are used to analyze system behavior under different lighting conditions and gesture variations.

A. System Workflow

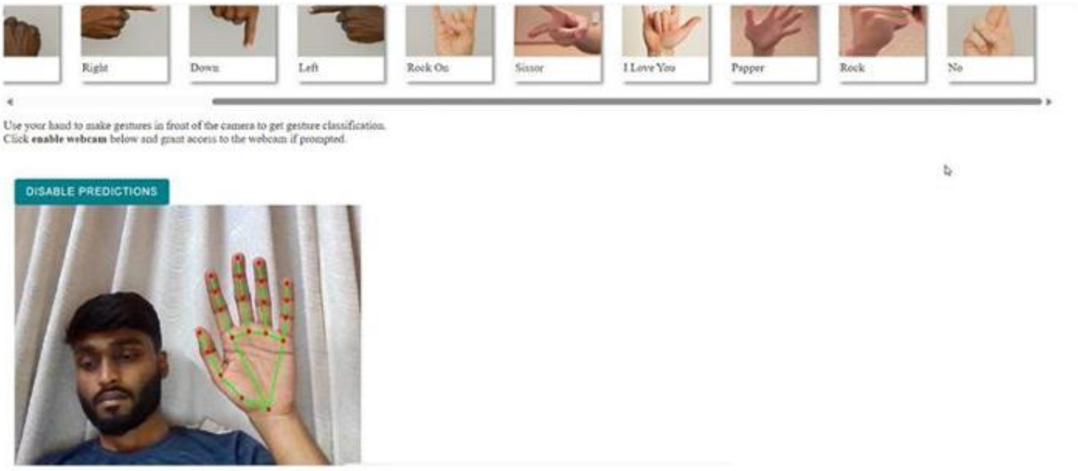
The workflow integrates gesture capture, preprocessing, CNN-based classification, and voice synthesis. Each stage is optimized to minimize latency and ensure real-time performance.

B. Simulation Setup

Various gestures are tested using different hand orientations and distances from the camera. The system is evaluated for consistent recognition and error handling.

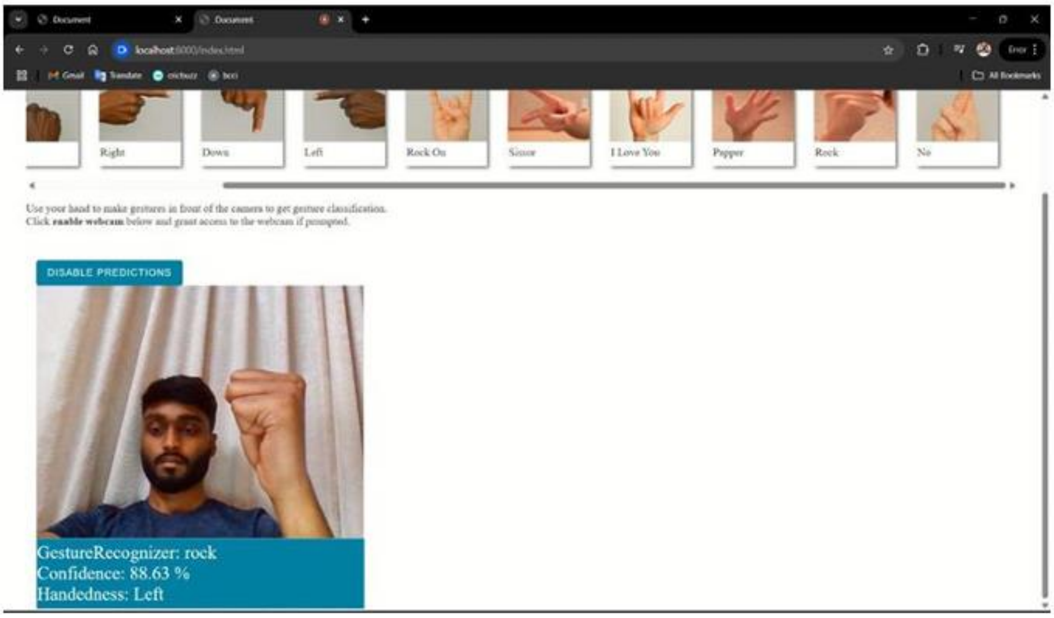
c. Results and Observations

- High gesture recognition accuracy was achieved
- Real-time voice output with minimal delay
- Stable performance across repeated trial



Use your hand to make gestures in front of the camera to get gesture classification.
Click **enable webcam** below and grant access to the webcam if prompted.

Fig 7.2.1 Gesture recognition



Use your hand to make gestures in front of the camera to get gesture classification.
Click **enable webcam** below and grant access to the webcam if prompted.

GestureRecognizer: rock
Confidence: 88.63 %
Handedness: Left

DOI: 10.17148/IJARCCE.2026.151138

Model Performance and Adaptability Analysis

The gesture recognition model demonstrated stable and consistent performance during training and testing phases. The CNN achieved reliable classification accuracy with smooth convergence and minimal loss variation. Real-time testing showed low response latency, enabling quick and accurate conversion of gestures into synthesized voice output.

V. RESULTS AND DISCUSSION

The Gesture Recognition for Voice Synthesis system successfully converts hand gestures into audible speech. The CNN model demonstrates reliable accuracy, while the voice synthesis module provides clear and understandable output. The system performs efficiently in real-time scenarios, making it suitable for assistive communication and human–computer interaction applications.

VI. CONCLUSION

This paper presented a real-time Gesture Recognition for Voice Synthesis system using computer vision and deep learning techniques. The integration of CNN-based gesture classification with text-to-speech conversion provides an effective communication solution for speech-impaired individuals. Experimental results confirm the system's accuracy, reliability, and usability..

VII. FUTURE WORK

Future enhancements include supporting dynamic gestures, multilingual voice synthesis, mobile deployment, and integration with wearable devices. The system can also be extended using advanced deep learning architectures and 3D gesture recognition.

REFERENCES

[1]. Priyakanth R., *Hand Gesture Recognition and Voice Conversion for Speech Impaired*, ResearchGate, 2021. Available: https://www.researchgate.net/publication/347242673_Hand_Gesture_Recognition_and_Voice_Conversion_for_Speech_Impaired

[2]. Brandone Fonya, *Real-Time Sign Language Gestures to Speech*, arXiv, 2025. Available: https://arxiv.org/abs/2508.12713

[3]. K. Patil, S. Ladake, S. Nirgude, V. Naphade, and P. Thummalakunta, "Translating Hands Gestures into Text and Speech," *International Journal of Innovative Research in Interdisciplinary Studies*, 2025. Available: https://www.ijirid.in/4-1-25Feb/4-1-16-Kunika%20Patil-Sakshi%20Ladake-Shraddha%20Nirgude-Vaishnavi%20Naphade-Praveen%20Thummalakunta.pdf

[4]. *Hand Gesture Recognition and Speech Synthesis Framework*, *International Journal of Scientific Research in Engineering and Management*, 2025. Available: https://ijsrem.com/download/hand-gesture-recognition-and-speech-synthesis-framework

[5]. *Real Time Hand Gesture to Speech Conversion*, *International Research Journal of Modern Engineering and Technology Solutions*, 2024. Available: https://www.irjmets.com/upload_newfiles/irjmets71100105732/paper_file/irjmets71100105732.pdf

[6]. "Gesture Language Translation Using Convolutional Neural Network," *Deep Learning Sign Language Translator*, 2025. Available: https://colab.ws/articles/10.1109%2Fesci63694.2025.10988297

[7]. Jinsu Kunjumon and R. K. Megalingam, *Sign Language to Speech Translation*, Semantic Scholar repository, 2024. Available: https://www.semanticscholar.org/paper/Sign-Language-to-Speech-Translation-Sharma-Panda/da3da732bd3194a52fceb55eafa995dc94b8672d

[8]. A Review Article on Deep Learning for Sign Language Recognition, *TechScience Press*, 2025. Available: https://www.techscience.com/CMES/v139n3/55626/html

[9]. "A Review of the Hand Gesture Recognition System: Current Progress and Future Directions," 2021. Available: https://www.scribd.com/document/795293236/A-Review-of-the-Hand-Gesture-Recognition-IEEE-2021

[10]. A Hand Sign Recognition-Based Communication System for Mute Individuals, *ScienceDirect*, 2025. Available: https://www.sciencedirect.com/science/article/pii/S221501612500514X