Journal of Computer and Electrical Sciences (JCES)- Volume-1 Issue-4, April 2026
ISSN: 3049-2602

HOLOSIGN: UNIVERSAL GESTURE COMMUNICATION SYSTEM

[1] Y. Karuna Manjusha, [2] S. Keerthi
[1] Assistant Professor, Department of Computer Science and Engineering
[2] Student, Department of Computer Science and Engineering
Andhra Loyola Institute of Engineering and Technology
Email: karunamanjusha@aliet.ac.in
Email: keerthisingamsetty093@gmail.com

Abstract: *Communication barriers for hearing-impaired individuals necessitate effective assistive solutions. This paper presents HoloSign, a multimodal communication system integrating gesture recognition, text-to-speech (TTS), and speech-to-text (STT) functionalities. The gesture recognition module uses a real-time camera and a trained machine learning model to identify sign language gestures through feature extraction and classification. The TTS module converts text input into speech using the Web Speech API, while the STT module transcribes spoken language into text via speech recognition APIs. The system is designed as an integrated application ensuring real-time performance, usability, and accessibility. Experimental results indicate reliable gesture recognition and efficient speech processing. HoloSign enhances communication by providing a unified platform that bridges the gap between hearing-impaired individuals and the general population.*

Keywords: Assistive Technology, Sign Language Recognition, Computer Vision, Machine Learning, Speech-to-Text, Text-to-Speech, Accessibility

I. INTRODUCTION

Communication barriers between hearing-impaired individuals and the general population pose significant challenges in daily interactions. Traditional sign language requires mutual understanding, which limits accessibility for non-sign-language users. With advancements in computer vision and machine learning, automated sign language recognition systems have emerged as effective assistive technologies. This paper presents HoloSign, a real-time multimodal communication system that integrates gesture recognition with speech processing. The system leverages MediaPipe for hand landmark detection and Scikit-learn for gesture classification. A Streamlit-based web interface enables interactive, user-friendly communication. The objective is to provide an efficient, low-cost, and scalable solution for inclusive human–computer interaction.

II. LITERATURE REVIEW

Sign language recognition and assistive communication systems have progressed from basic image processing methods to advanced deep learning and multimodal frameworks. Earlier approaches lacked robustness, while modern techniques emphasise real-time performance, accuracy, and usability. Additionally, speech processing technologies have enhanced communication capabilities. However, most existing works focus on single-modality systems, highlighting the need for integrated solutions like HoloSign.

- Early vision-based methods using skin segmentation and contour detection suffered from sensitivity to lighting and background variations.
- Machine learning models such as Support Vector Machines (SVM) and Random Forests improved classification but relied on handcrafted features.
- Deep learning approaches, particularly Convolutional Neural Networks (CNNs), enabled automatic feature extraction and higher accuracy in gesture recognition.

- Frameworks like MediaPipe introduced efficient real-time hand tracking, improving system responsiveness and usability.
- Speech processing technologies, including Speech-to-Text (STT) and Text-to-Speech (TTS), enhanced bidirectional communication.
- Existing systems are predominantly unimodal, lacking integration of gesture and speech, which limits overall communication effectiveness.

III. PROPOSED SYSTEM

The proposed system, HoloSign, is a multimodal assistive communication platform designed to enable seamless interaction between hearing-impaired individuals and the general population. The system integrates three primary modules: gesture recognition, text-to-speech (TTS), and speech-to-text (STT). These modules operate within a unified interface to provide real-time, bidirectional communication. In the gesture recognition module, video input is captured through a camera and processed using hand landmark detection techniques. Extracted features are passed to a trained machine learning model for gesture classification, and the output is displayed as text. The text module converts user-input text into speech using TTS, while the voice module converts spoken input into text using STT. All components are integrated using a Streamlit-based interface to ensure usability and efficiency.

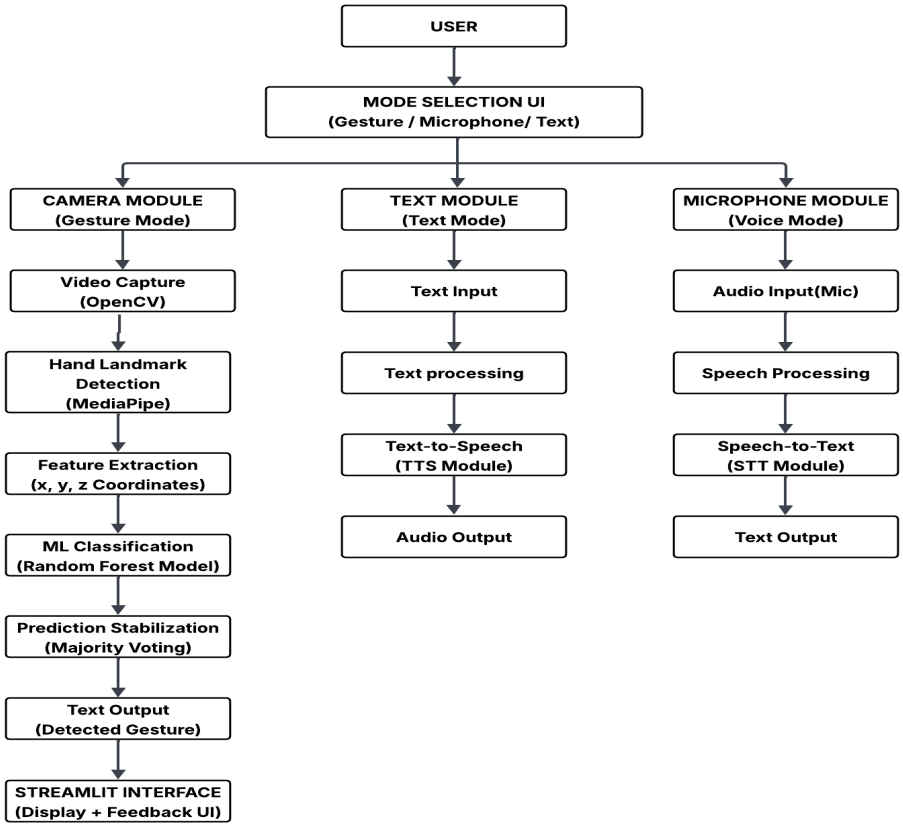

Fig. 1. System Architecture of HoloSign

The system architecture illustrates the flow of data from user input through the respective modules. The gesture module processes visual data using OpenCV and MediaPipe for feature extraction and classification

via a Random Forest model. The text module performs text processing followed by speech synthesis using the Web Speech API. Similarly, the voice module captures audio input and converts it into text using speech recognition APIs. The outputs from all modules are displayed through a unified interface, ensuring real-time feedback and improved accessibility.

IV. METHODOLOGY

The HoloSign system operates through a multimodal pipeline where user-selected inputs (gesture, text, or speech) are processed using computer vision and speech APIs to generate corresponding outputs in real time. The system integrates gesture recognition via a machine learning model with text-to-speech and speech-to-text modules to enable seamless and accessible communication.

The methodology of the system is organised into the following steps:

1. Mode Selection:

The user selects the desired mode (Gesture Mode / Text Mode / Voice Mode) through the Streamlit-based user interface. The system activates the corresponding module based on the selection.

2. Gesture Data Acquisition:

In gesture mode, the camera is activated using OpenCV to capture real-time video frames. Continuous frame acquisition ensures smooth and dynamic gesture detection.

3. Hand Landmark Detection:

Each captured frame is processed using MediaPipe to detect hand landmarks. The system extracts key points representing hand structure and motion.

4. Feature Extraction:

The detected landmarks are converted into numerical features in the form of (z)coordinates. These features serve as input for the classification model.

5. Gesture Classification:

The extracted features are passed to a trained Random Forest model, which predicts the corresponding gesture. Majority voting is applied across multiple frames to improve prediction stability.

6. Text-to-Speech Processing:

In text mode, user input text is processed using the Web Speech API. The system converts the text into audible speech output for communication.

7. Speech-to-Text Processing:

In voice mode, audio input is captured through a microphone. Speech recognition APIs convert spoken language into text for user understanding.

8. Output Display and Feedback:

The system displays results in real time through the Streamlit interface, including detected gestures, converted text, or generated speech output.

9. Data Storage:

Recognised gestures and outputs are stored in a history module for future reference and analysis.

10. System Integration and Control:

All modules are integrated into a unified application, ensuring smooth switching between modes, real-time processing, and efficient user interaction.

V. PROPOSED SYSTEM RESULTS

The proposed HoloSign system was successfully developed and tested under real-time conditions. The system effectively performed gesture recognition, text-to-speech (TTS), and speech-to-text (STT) operations, enabling seamless multimodal communication.

- The gesture recognition module accurately identified 10 predefined gestures, namely: *Hello, Thank You, Yes, No, Help, Sorry, Alright, Friends, Happy,* and *How Are You*.
- The system achieved stable predictions with confidence scores typically above 50%, ensuring reliable output.
- MediaPipe-based hand landmark detection enabled precise feature extraction, improving classification accuracy.
- The Random Forest model demonstrated consistent performance across varying lighting and background conditions.
- Majority voting across frames reduced prediction fluctuations and enhanced stability.
- The text-to-speech module successfully converted input text into clear audio output.
- The speech-to-text module accurately transcribes spoken input into text with minimal delay.

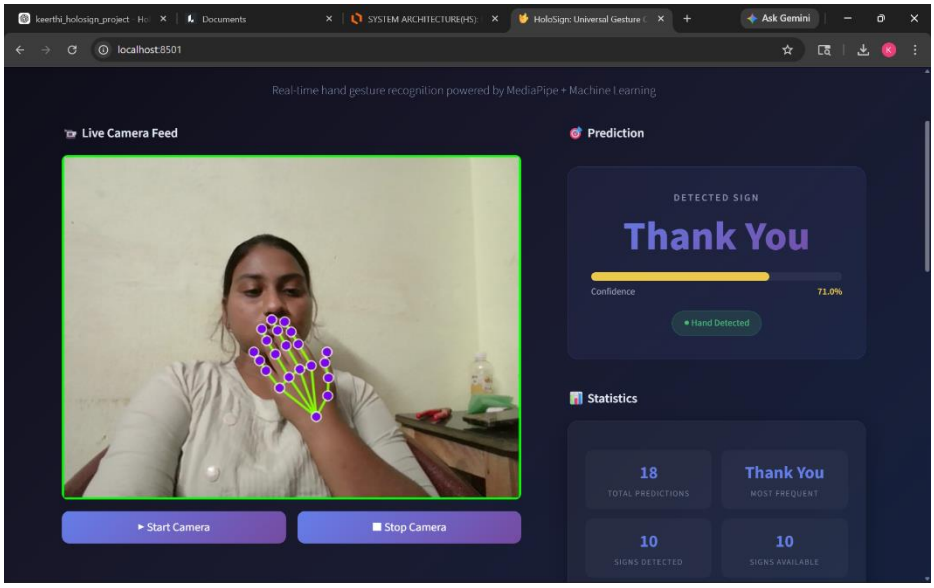

Fig. 2. Gesture Recognition Output Interface

This interface shows the real-time hand gesture recognition system using MediaPipe and machine learning, where the camera captures hand movements and predicts the corresponding sign with a confidence score.

The system is capable of detecting 10 predefined gestures—Hello, Thank You, Yes, No, Help, Sorry, Alright, Friends, Happy, and How Are You—as reflected in the statistics panel.

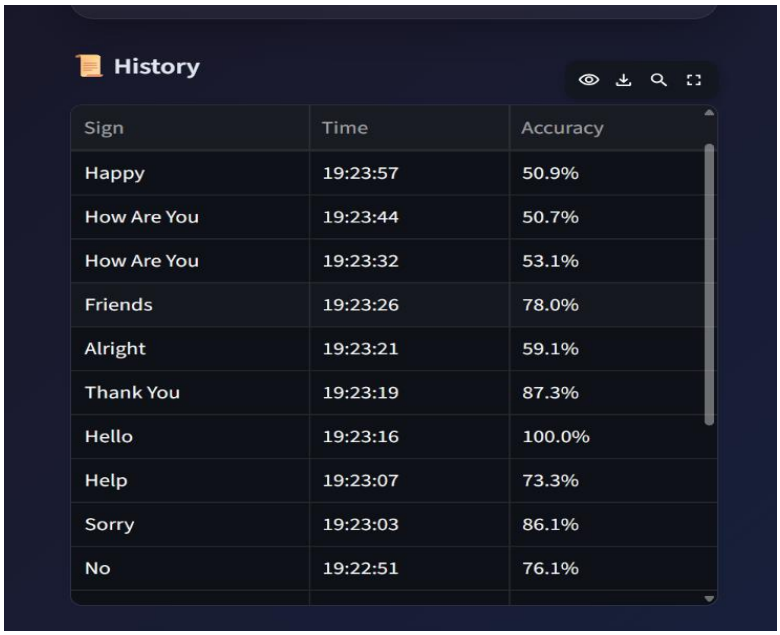

Fig. 3. History Table for Gesture Recognition

The History Table (Fig. 3) stores all previously recognised gestures along with their predicted labels and timestamps, enabling users to track and review past interactions for better analysis.

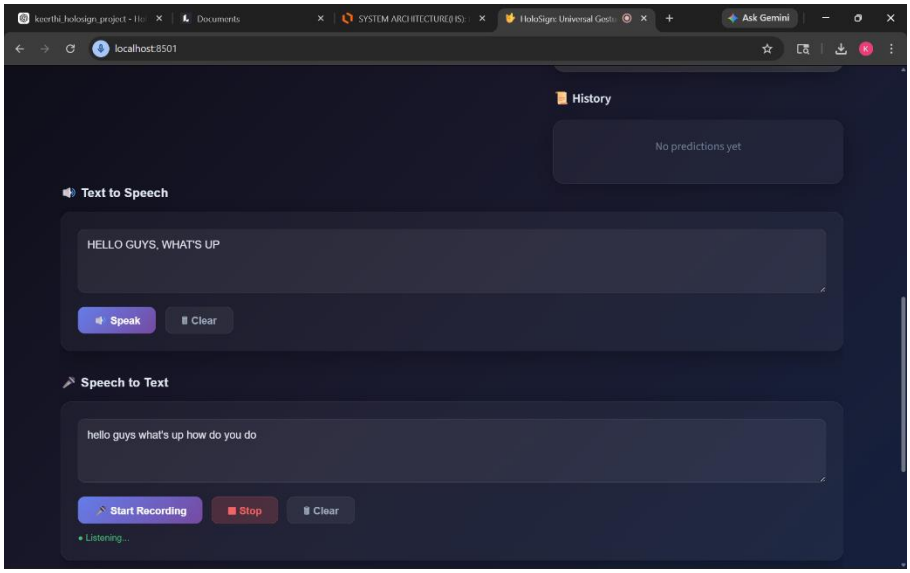

Fig. 4. Text-to-Speech and Speech -to-Text Modes

In Fig. 4, the Text-to-Speech module converts user-entered text into audible speech using the Web Speech API, while the Speech-to-Text module captures spoken input and accurately transcribes it into text for seamless communication.

Overall, the system exhibited fast response time, reliable gesture recognition, and accurate recognition of 10 essential gestures, fast response time, and efficient multimodal communication, making it a reliable assistive solution for real-world use. and efficient speech processing, demonstrating its effectiveness as an assistive communication tool. The integration of multiple modules into a single platform significantly improves accessibility and real-time communication for hearing-impaired users.

VI. CONCLUSION

This work successfully developed **HoloSign**, a low-cost and efficient multimodal communication system designed to assist hearing-impaired individuals. The system integrates gesture recognition, speech-to-text (STT), and text-to-speech (TTS) functionalities within a unified platform. The gesture recognition module effectively identified 10 predefined hand gestures in real time using MediaPipe-based feature extraction and a Random Forest classifier, ensuring reliable and stable predictions. The implementation of speech processing modules using Web Speech APIs enabled accurate and fast conversion between text and speech, facilitating seamless bidirectional communication. The use of a Streamlit interface provided an interactive and user-friendly environment for real-time input and output visualisation. Overall, the system demonstrated efficient performance, low latency, and high usability, making it a practical assistive solution. By integrating multiple communication modes into a single platform, HoloSign enhances accessibility and reduces communication barriers for the hearing-impaired community.

REFERENCES

1. *F. Chollet, Deep Learning with Python, Manning Publications, 2018.*
2. *I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.*
3. *Google, "MediaPipe: Framework for Building Perception Pipelines," 2023.*
4. *OpenCV, "Open Source Computer Vision Library Documentation," 2024.*
5. *Mozilla, "Web Speech API Documentation," 2023.*
6. *T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning, Springer, 2009.*
7. *L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.*
8. *S. Mitra and T. Acharya, "Gesture Recognition: A Survey," IEEE Transactions on Systems, Man, and Cybernetics, 2007.*
9. *R. Kadous, "Machine Recognition of Auslan Signs Using PowerGloves," Proceedings of ICSLP, 1996.*
10. *K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," 2015.*
11. *D. Pavllo et al., "3D Hand Pose Estimation in the Wild," CVPR, 2019.*
12. *Streamlit, "Streamlit Documentation: Build Data Apps," 2024.*