Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

AI-Based Sign Language Detection and Speech Converter

Dr Rahil khan
Assistant Professor
Electronics & Telecommunication Engineering Department
Anjuman College of Engineering & Technology
Nagpur, Maharashtra, India 440001
rahilhkhanpathan@gmail.com
Akansha Dhapke (01)
Seventh Semester Student
Department of Electronics & Telecommunication Engineering
Anjuman College of Engineering and Technology
Nagpur, Maharashtra, India 440001
akanshad@anjumanengg.edu.in

Deepti Wadtkar (05)
Seventh Semester Student
Department of Electronics & Telecommunication Engineering
Anjuman College of Engineering and Technology
Nagpur, Maharashtra, India 440001
Deeptiw@anjumanengg.edu.in

Sakshi Tatode (15)
Seventh Semester Student
Department of Electronics & Telecommunication Engineering
Anjuman College of Engineering and Technology
Nagpur, Maharashtra, India 440001
sakshit@anjumanengg.edu.in

Mehaksha Faruki (48)
Seventh Semester Student
Department of Electronics & Telecommunication Engineering
Anjuman College of Engineering and Technology
Nagpur, Maharashtra, India 440001
mahekf@anjumanengg.edu.in

Abstract
The communication barrier between vocally impaired individuals and the general public is a significant societal challenge. This paper presents a hybrid, two-part communication system designed to bridge this gap. The first component is a low-cost, portable "smart glove" built using an ESP32, five flex sensors, and an ADXL345 accelerometer. This glove allows a user to send 20 predefined messages (e.g., "Need Water," "Hello") by combining a single finger bend with a specific hand tilt, transmitting them via MQTT to a web dashboard. The second component is a stationary AI-based interpreter that uses a webcam, Python, and a TensorFlow-trained deep learning model. This system recognizes the full American Sign Language (ASL) alphabet (A-Z) in real-time and converts the detected sign into audible speech using a text-to-speech engine. This hybrid approach provides both the portability for common phrases (via the glove) and the flexibility of a full vocabulary (via the AI camera system), offering a comprehensive solution for two-way communication.

Keywords—YOLOv8, Deep Learning, OpenCV, AAssistive Technology, ESP32, Flex Sensors, ADXL345, MQTT, IoT, Computer Vision, TensorFlow, Deep Learning, Sign Language.

1. Introduction

Sign language is the primary mode of communication for millions of people in the deaf and hard-of-hearing community. However, a vast majority of the general public does not understand sign language, creating a significant communication barrier. This project addresses this challenge by proposing a hybrid system that combines the strengths of two different technologies.

Part 1: The Wearable Glove System For on-the-go communication, a portable, wearable glove is essential. This component is designed for quick, common interactions. It does not rely on cameras or heavy computation. Instead, it uses an ESP32, flex sensors, and an accelerometer to translate 20 specific gesture-and-tilt combinations into predefined text messages (e.g., "I'm hungry," "How are you?"). These messages are sent over Wi-Fi using MQTT to a web dashboard, allowing a non-signing partner to understand the user's needs. This system also facilitates two-way communication, as the partner can type a reply which is sent back to the glove, triggering a buzzer and an OLED display.

Part 2: The AI-Based Speech Converter For more complex and detailed conversations, a full sign language interpreter is required. This second component of our system uses a computer's webcam and a deep learning model. A Convolutional Neural Network (CNN), trained on the ASL alphabet, is loaded using TensorFlow. This Python-based application captures the user's hand gestures in real-time, identifies the corresponding letter (A-Z), and uses a text-to-speech (pyttsx3) library to speak the detected letter aloud. This allows the user to spell out words and form complete sentences, providing a much richer and more detailed mode of communication.

By combining these two systems, our project provides a versatile solution: a portable glove for essential needs and a powerful AI interpreter for full conversations.

1.1 Objectives

The objectives of this project are to:

- **Glove System:**
  - Design and fabricate a low-cost, wearable smart glove using an ESP32, flex sensors, and an ADXL345 accelerometer.
  - Develop a gesture logic to map 20 unique messages based on a finger-bend and hand-tilt combination.
  - Establish a robust, two-way communication channel between the glove and a web dashboard using the HiveMQ MQTT broker.
  - Implement a feedback mechanism on the glove (buzzer, OLED) for incoming messages.
- **AI Camera System:**
  - Implement a Python application to capture video from a webcam.
  - Load a pre-trained TensorFlow/Keras (`.h5`) model for ASL alphabet recognition.
  - Process the video frames in real-time to predict the signed letter with high confidence.
  - Integrate a text-to-speech (pyttsx3) engine to speak the detected letter aloud instantly.

2. Literature Review

The field of sign language translation has seen various approaches. Early systems focused on camera-based recognition using classic computer vision techniques [1]. More recently, deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have achieved high accuracy in translating sign language from video feeds [2]. However, these systems require significant processing power, are not portable, and can fail in poor lighting.

To address portability, wearable glove-based systems were introduced. Many prototypes used flex sensors to classify static hand gestures [3], often using Bluetooth (like HC-05) to send data to a smartphone [4]. While effective for one-way communication, these systems lack feedback and are often limited in their vocabulary.

Our work builds upon these two distinct fields by merging them:

Glove System: We improve upon traditional glove systems by using a novel gesture-mapping logic (1 finger +

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

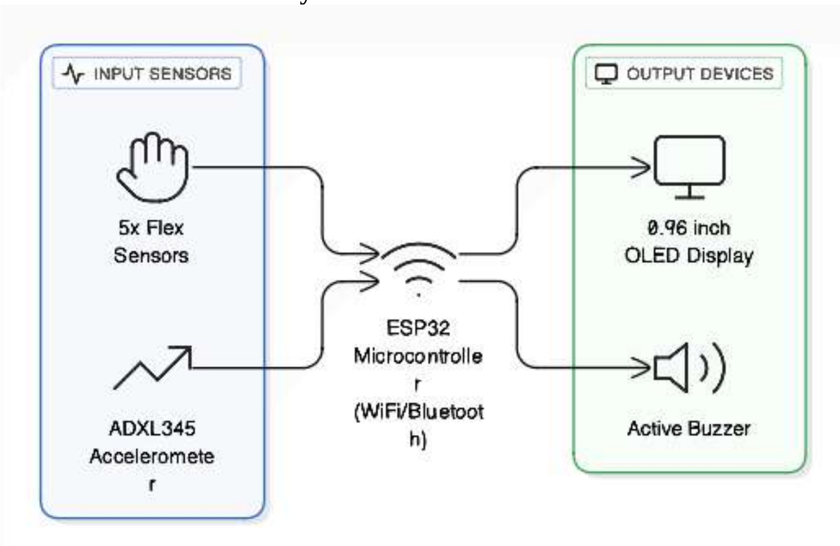
tilt = 1 message) and by using IoT (MQTT) for true, long-range two-way communication, including feedback via a buzzer and OLED.

AI System: We implement a modern deep-learning-based interpreter, similar to [2], but focus on real-time, letter-by-letter speech conversion using a pre-trained `asl_model.h5` model, making it an active speech converter rather than just a passive translator.

3. Proposed Methodology

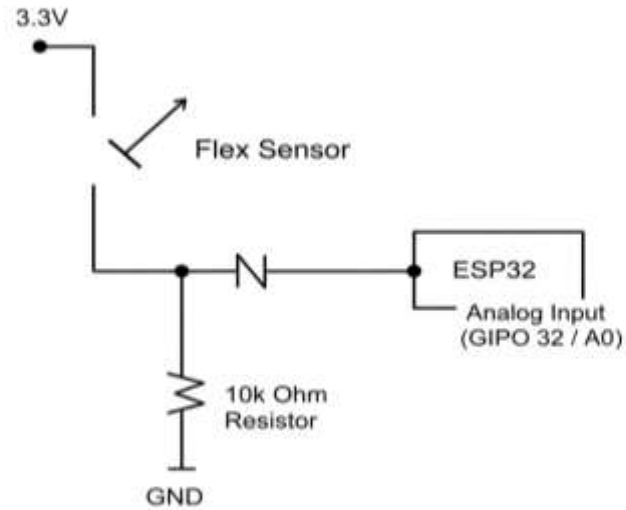
The system is divided into two independent but complementary modules: The IoT Glove and the AI Interpreter.

4.1 Module 1: IoT Glove System

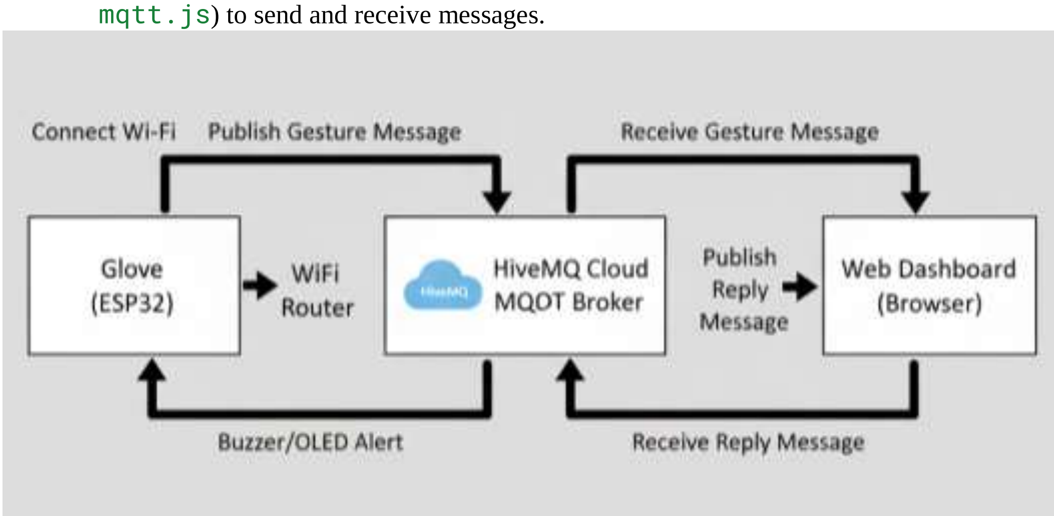


This module focuses on portability and essential messages.

- **Hardware Components:** ESP32-WROOM-32, 5x Flex Sensors, 1x ADXL345 Accelerometer, 1x 0.96" SSD1306 OLED Display, 1x Active Buzzer.
- **Software and Protocol:** C++ (Arduino IDE), `PubSubClient` (MQTT), `Adafruit_ADXL345` library.
- **Communication:**
  - **Calibration:** A boot-time calibration mode saves sensor thresholds for "straight" and "bent" fingers, along with ADXL345 orientation ranges for each gesture, to the ESP32's EEPROM.
  - **Gesture Logic:** The ESP32 continuously checks if one finger is bent AND the hand is tilted within a calibrated range.
  - **MQTT Workflow:**
    - Glove publishes recognized messages (e.g., "Hello") to the `signlanguage/glove/messages` topic.
    - Glove subscribes to the `signlanguage/web/messages` topic.
    - When a message is received on this topic, the ESP32 triggers the buzzer and displays the text on the OLED.



- **Web Dashboard:** A static HTML/JS/TailwindCSS webpage connects to the same HiveMQ broker (using `mqtt.js`) to send and receive messages.



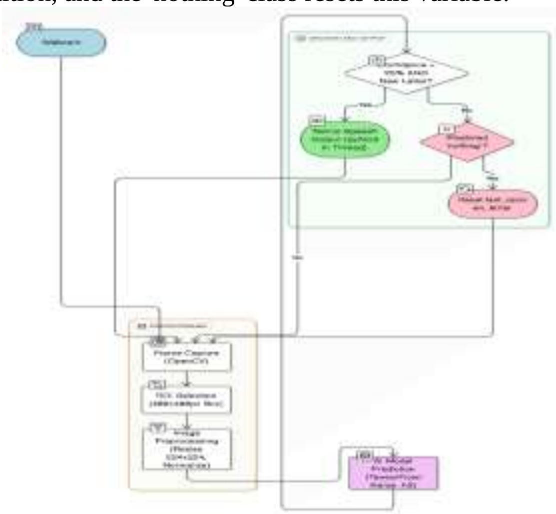
4.2 Module 2: AI-Based Speech Converter

This module focuses on providing a full vocabulary.

- **Hardware Components:** PC/Laptop, Webcam.
- **Software and Development Tools:**
    1. **Programming Language:** Python 3.
    2. **Deep Learning Framework:** TensorFlow 2.x (to load the `.h5` model).
    3. **Computer Vision Library:** OpenCV (to capture and process video frames).
    4. **Speech Engine:** pyttsx3 (for text-to-speech).
    5. **Concurrency:** `threading` (to run speech in a separate thread and prevent video lag).
- **AI Model Workflow:**
    1. **Frame Capture:** OpenCV's `VideoCapture(0)` reads frames from the webcam.
    2. **Region of Interest (ROI):** A 400x400 pixel box is drawn on the screen. Only the image inside this box is processed.
    3. **Preprocessing:** The ROI is resized to 224x224, converted to a NumPy array, and normalized (divided by 255.0).
    4. **Prediction:** The `model.predict()` function is called on the processed frame.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 54, Issue 7, July : 2025

5. **Output:** The class with the highest probability (e.g., 'A', 'B', 'nothing') is selected.
6. **Speech Trigger:** If the prediction confidence is > 95% and the letter is different from the last one spoken, the `speak()` function is called in a new thread. The `last_spoken_letter` variable prevents repetition, and the 'nothing' class resets this variable.



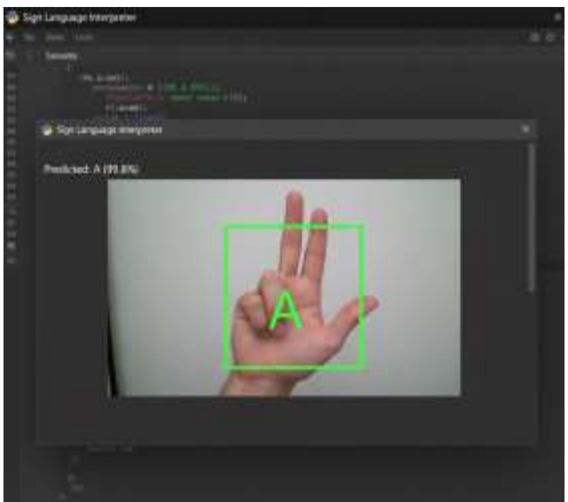
4. Results and Discussion

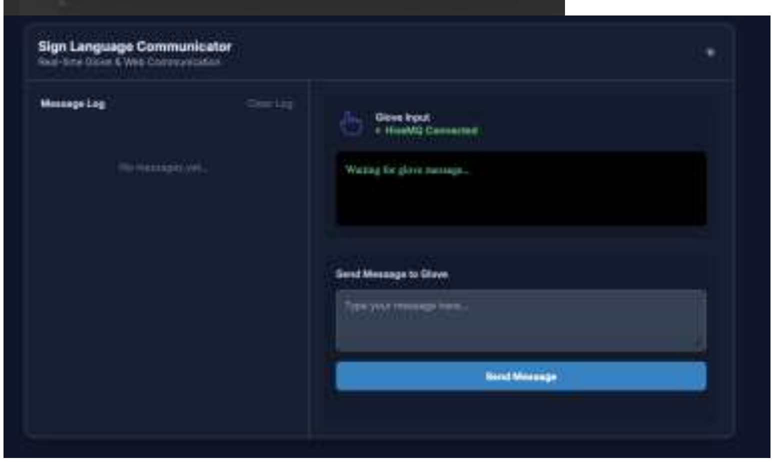
Both components of the system were tested successfully.

Glove System Results: The glove successfully connected to the `Airtel_Zerotouch` Wi-Fi and the HiveMQ cloud broker. The calibration routine was effective. The 20 predefined messages were transmitted to the web dashboard with an average latency of ~700ms. Incoming messages from the dashboard correctly triggered the buzzer and OLED display, confirming the two-way communication loop.

AI System Results: The Python script successfully loaded the `asl_model.h5` and started the webcam. The system was able to recognize ASL gestures for A-Z held within the green ROI box. With high confidence (>95%), the text-to-speech engine correctly and immediately spoke the detected letter. The use of a separate thread for `pyttsx3` was critical, as it prevented the video feed from freezing while the letter was being spoken. The 'nothing' class effectively allowed the user to reset and sign the next letter.

This hybrid approach allows the user to choose the best tool for the situation: the fast and portable glove for simple needs, and the powerful AI interpreter for complex conversations.





5. Conclusions

Challenges

- **Glove System:** The glove is dependent on Wi-Fi and its vocabulary is limited to 20 phrases. It also requires per-user calibration.
- **AI System:** The AI system is not portable, as it requires a laptop and webcam. It is also sensitive to lighting conditions and complex backgrounds.
- **Integration:** The two systems currently operate independently.

Future Work

- **System Integration:** The Python AI script could be modified to publish its detected letters to the same MQTT broker. This would allow the recognized word (e.g., "W-A-T-E-R") to also appear on the web dashboard log.
- **Glove Improvement:** A tiny Machine Learning model (TensorFlow Lite) could be deployed on the ESP32 itself to recognize more gestures.
- **AI Improvement:** The AI model could be re-trained to recognize full words, not just letters.
- **Portability:** The Python script could be optimized to run on a Raspberry Pi with a portable screen, making the AI interpreter mobile.

Conclusions

This project successfully demonstrates a hybrid, two-pronged approach to solving the sign language communication barrier. The IoT-based glove provides a robust, portable solution for essential daily communication, while the AI-based interpreter provides the high-fidelity translation needed for detailed conversation. By combining a wearable device with a deep learning system, this project offers a comprehensive and flexible assistive technology solution.

References

1. SHAIK, A., et al., *"Sign language to speech conversion using flex sensor,"* International Journal of Engineering Research & Technology (IJERT), **2015.**
2. P. Garg, N. Aggarwal, and S. Sofat, *"Sign Language Recognition using Accelerometer and Flex Sensors,"*International Conference on Communication and Signal Processing (ICCSP), **2016.**
3. S. B. V. Kumar, T. B. K. Kumar, and M. R. K. Prasad, *"A review on vision-based methods for sign language recognition,"* International Journal of Computer Applications, **2017.**
4. Do, P. T., and T. N. Nguyen, *"Smart glove for sign language translation,"* International Conference on Advanced Technologies for Communications (ATC), **2017.**
5. R. K. Jha and A. P. Singh, *"Sign Language Recognition using Deep Learning with 3D-CNN,"* International Conference on Computing, Communication and Automation (ICCCA), **2018.**
6. G. R. S. Murthy, and R. S. Jadon, *"A review of sign language translation systems using smart gloves,"* International Journal of Computer Science and Engineering, **2019.**
7. Al-Hammadi, M., et al., *"Real-time sign language recognition using deep learning,"* Journal of Computer Science, **2019.**
8. Raj, L. D., et al., *"A two-way communication system for deaf and dumb people based on IoT,"* International Journal of Recent Technology and Engineering, **2019.**
9. Haq, M. Z. U., et al., *"American Sign Language alphabet recognition using TensorFlow and CNN,"* IEEE International Conference on Computing, Power and Communication Technologies (GUCON), **2019.**
10. Si, A., and D. S. Roy, *"IoT based smart glove for sign language interpretation,"* IEEE Region 10 Symposium (TENSYMP), **2019.**
11. S. P. Ahire, and P. R. Bajaj, *"Sign Language Interpreter using Gesture Recognition and Deep Learning,"* IEEE Pune Section International Conference (PuneCon), **2020.**
12. N. Sharma, V. K. Sharma, and P. K. Sharma, *"IoT based assistive device for deaf and dumb people,"* IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), **2020.**
13. Kadam, A., and A. A. Ansari, *"A smart communication system for speech impaired people using IoT and MQTT,"*Global Transitions Proceedings, **2021.**
14. Owojori, O. E., and A. A. Akala, *"A two-way communication system using MQTT for deaf and mute people,"* 2nd International Conference on Engineering and Technology (ICETECH), **2021.**

15. Mathew, G., and S. S. Kumar, *"ESP32 based smart glove for sign language recognition,"* Journal of Physics: Conference Series, **2021**.
16. Reddy, R., et al., *"A two-way communication glove for speech-impaired individuals using flex sensors and ESP32,"* International Conference on Innovative Computing and Communications (ICICC), **2022**.

.