RESEARCH ARTICLE OPEN ACCESS

VISION BASED DEEP LEARNING FRAMEWORK FOR SIGN LANGUAGE CONVERSION

S. Kumaran[1], S.G. Nandhini[2] A. Ushananthini[3]

G. Jayanthini[4]
Students[1,2,3] Assistant Professor[4]
Department Of ECE

Parisutham Institute of Technology and Science, Thanjavur
[1]kumaranshanmugam1021@gmail.com ; [2]jayanthi070584@gmail.com

ABSTRACT: ***This paper introduces a real-time, deep learning framework designed to bridge communication gaps for individuals with hearing and speech impairments. By leveraging the YOLOv8 object detection algorithm, the system accurately identifies American Sign Language (ASL) gestures from live video feeds.***

To balance portability with performance, the framework uses a hybrid edge-cloud architecture:

- ***Edge: A Raspberry Pi Zero 2 W equipped with a Pi Camera captures image data.***
- ***Cloud: Data is transmitted to an AWS EC2 instance where a FastAPI backend manages the inference process.***

The system demonstrated high reliability, achieving a mean Average Precision (mAP@0.5) of 95.6%. By instantly converting hand signals into readable text, this solution offers a cost-effective and portable tool for enhancing accessibility in healthcare, education, and public sectors.

KEYWORDS: YOLOv8, Sign Language Recognition, ASL, Deep Learning, Computer Vision, Raspberry Pi, AWS, FastAPI, Edge-cloud Architecture.

INTRODUCTION:

Effective communication is a fundamental part of daily life, yet millions of people with hearing and speech impairments face significant barriers every day. Globally, approximately 466 million people—including over 18 million in India—rely on sign language as their primary means of expression. Because the general public often lacks sign language skills, a persistent "communication gap" exists in critical settings like hospitals, government offices, and public service centres.

Past attempts to bridge this gap have often fallen short. Human interpreters are expensive and not always available, while text-based tools can be slow. Even specialized hardware, such as sensor-equipped gloves, can be bulky, uncomfortable, and impractical for long- term use.

Fortunately, recent breakthroughs in **computer vision** and **deep learning** are opening new doors. We can now use standard cameras to recognize and interpret hand gestures with incredible speed. By using **YOLOv8**, a cutting-edge object detection model, we can build a system that is not only highly accurate but fast enough to translate sign language into text in real time. This technology offers a portable, high-performance solution for making the world more accessible and inclusive.

EXISTING METHOD:

Historically, bridging the communication gap has relied on a mix of hardware and early-stage software. Some systems used **specialized sensors or wearable gloves** to track hand movements physically. In the realm of computer vision, many developers turned to **MediaPipe** for basic hand-tracking or utilized older deep learning architectures, such as **YOLOv5**. These tools were designed to identify hand shapes and classify them into specific letters or words. While these setups proved successful in "laboratory" settings—with perfect lighting and simple backgrounds—they often struggled when moved into the unpredictable nature of the real world.

DRAWBACK:

Despite their contributions, these existing methods face several significant hurdles:

- **Physical Constraints:** Sensor-based wearables are often expensive, fragile, and uncomfortable for the user to wear throughout a typical day.
- **Environmental Sensitivity:** Software like MediaPipe frequently loses accuracy in low-light conditions or busy visual environments.
- **Performance Lags:** Older AI models are often too slow for natural conversation, leading to "lag" that makes real-time interaction frustrating.
- **Complexity Issues:** Many current systems struggle to distinguish between similar-looking gestures or more complex signs, leading to frequent translation errors.
- **Lack of Portability:** Because many of these models require heavy processing power, they are rarely optimized for the mobile or "on-the-go" use that daily life demands.

motorbike unit (MU) have many sensors that ensure the safety of the rider.

PROPOSED METHOD:

A deep learning system based on vision is used in this research. For quick and precise gesture recognition, it makes use of YOLOv8. A camera on a Raspberry Pi is used to take pictures. A cloud server receives the pictures. The gesture is processed and recognized by the model. The outcome appears as text.

WORKING:

The illustration shows how an edge-cloud-based sign language recognition system operates. Initially, the camera takes real-time pictures of hand gestures. A processing unit (such as a Raspberry Pi) receives these pictures and uses the internet to send them to a cloud server. Before being sent to a deep learning model (YOLOv8) for gesture recognition and classification, the image is preprocessed on the server side (resized, normalized). The model recognizes the matching motion or letter in sign language. After then, the outcome is returned to the edge device. The identified output is then shown as text on a screen. Deaf and mute people can communicate in real time because to this ongoing procedure. A hybrid edge-cloud architecture is used by the suggested system to guarantee quick and precise sign language recognition. First, a Raspberry Pi Zero 2 W and Pi Camera Module 2 are used to record real-time hand gestures and send them via HTTP requests to an AWS server. This server serves as the main centre for complex computing an FastAPI backend. In our setup, the Raspberry Pi Zero 2 W serves as the primary controller. It takes pictures from the Pi Camera and processing them. For gesture recognition, it transmits the recorded data to the cloud server. It receives the outcome after processing and presents it as text. Additionally, it controls how hardware and software components communicate with one another. As a result, it permits the system to function in real time.

Images are preprocessed for scaling and normalization using OpenCV and NumPy upon arrival. A YOLOv8 model, which uses deep-learned features to identify and categorize motions, is then given these modified photos. The motion is recognized, translated into text, and sent back to the Raspberry Pi as a JSON response. Real-time text display is made possible by this effective procedure, offering a scalable and responsive communication tool

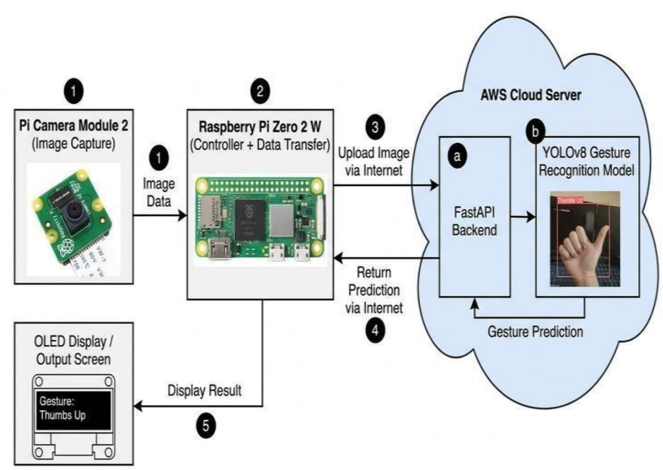
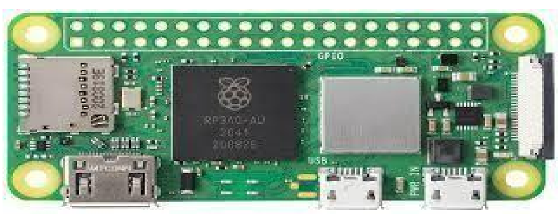


Figure: System Architecture of the Proposed Edge-Cloud Model

The project uses the Pi Camera Module 2 to take real-time pictures of hand movements. The Raspberry Pi is linked to an 8-megapixel camera. The camera continuously captures clear photos of gestures. The processing system receives these pictures in order to identify them. The deep learning model uses the recorded data to identify gestures. As a result, it is crucial for precise recognition of sign language.

In our setup, the Raspberry Pi Zero 2 W serves as the primary controller. It takes pictures from the Pi Camera and processing them. For gesture recognition, it transmits the recorded data to the cloud server. It receives the outcome after processing and presents it as text. Additionally, it controls how hardware and software components communicate with one another. As a result, it permits the system to function in real time.



In our setup, the OLED display serves as the output screen. In real time, it displays the identified sign language as text. The results of the processing are sent to the display by the Raspberry Pi. It gives consumers output that is easy to read and understand. This facilitates efficient communication between hearing and deaf individuals. As a result, it is crucial to showing the outcome.

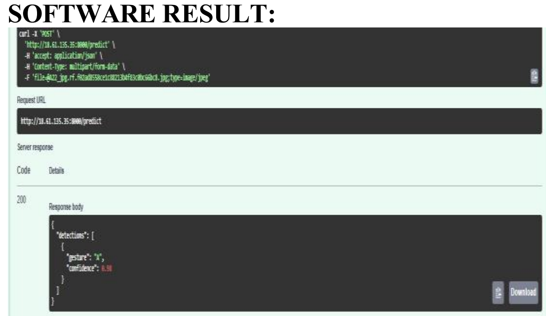
HARDWARE USED:

- Pi Camera Module 2
- Raspberry Pi Zero 2W
- LED Display
- MicroSD Card

SOFTWARE USED:

- YOLOv8
- AWS EC2 Instance
- FastAPI

SOFTWARE RESULT:



ADVANTAGES:

- Real-time recognition of sign language.
- High precision with deep learning (YOLOv8)
- Low-cost and portable system.
- Simple to use without the need for any devices.
- Enhances deaf and mute people's ability to communicate.
- Works in a variety of settings (background, lighting)
- Scalable using cloud computing lessens reliance on human interpreters.

APPLICATION:

- Utilized for patient communication in hospitals.
- Supports inclusive education in schools.
- Beneficial in government offices and banks.
- Suitable for usage in customer service, departments and hotels.
- Facilitates communication at work.
- Helpful in public areas such as airports and train stations.
- Can be used to help deaf and mute people communicate on a regular basis.

CONCLUSION:

Through the use of YOLOv8 and a hybrid edge-cloud architecture, this team successfully created a real-time Sign Language Recognition system. The Raspberry Pi Zero 2 W is used by the system to efficiently record hand motions, which are then processed by a cloud-based model for precise recognition. High precision, speed, and robustness under various environmental conditions are guaranteed by using YOLOv8. The solution lowers overall costs and hardware constraints by shifting processing to the cloud. It offers a practical way to translate sign language into text, enhancing deaf and mute people's ability to communicate. The study shows how embedded systems and deep learning may be used practically in assistive technology. For full-word recognition and practical use, it can be further improved.

REFERENCES:

Alsharif, M., Khan, A., and Ahmad, Z. (2024). YOLOv8-based ASL detection.
CNN-LSTM sign language system (Anturkar, A., Khot, A., and Megadump, A., 2025).
FastAPI for ML deployment, Bhatt, S., Trivedi, A., and Patel, R. (2023).
YOLOv8 framework, Jocher, G. et al., 2023.
3D CNN sign detection by Kumar, S., Rani, R., and Chaudhari, U. (2024).
Najib, F. M. (2025). Sign recognition in many languages. YOLOv11 sign recognition; Rahman, K., Islam, M., and Ahmed, S., 2025; Stipulator, M. et al., 2025. Deep learning for sign language proces