Edge AI for Real-Time Sign Language Translator

[1]Ravikumar Chawhan[1], Shivaraj Yaliwal[2], Akash P. Chandai[3], Iranna Jakkali[4], Puneeth Akki[5]
[1]Assistant Professor, [2–5]UG Students
Department of Information Science and Engineering
Smt. Kamala and Sri. Venkappa M. Agadi College of Engineering & Technology
Lakshmeshwar, Karnataka, India
ravismsk1@gmail.com

***Abstract*— Communicating with individuals who are deaf can often be difficult. This research aims to create a system that translates sign language into text and speech in real time. Using a camera, the system captures hand movements and converts them into written or spoken language. It relies on computer vision and a specialized computational model to analyze the hand gestures, interpret their meaning, and translate them accurately.**

All processing occurs directly on the device, ensuring fast performance and protecting user privacy, since data does not need to be sent to external servers. In tests, the system was able to correctly recognize about 90% of gestures, even under poor lighting or noisy conditions. This sign language translation system provides an effective tool to help people who are deaf communicate more easily with others.

Keywords: Convolutional Neural Networks (CNN), MediaPipe Hand Landmarks, Real-Time Translation, Computer Vision, Sign Language Recognition, Edge Computing, and Assistive Technology.

I. INTRODUCTION

Effective communication is crucial in daily interactions, but it can be particularly challenging for individuals who are deaf or hard of hearing when others do not understand sign language. Since sign language is not widely known, these individuals often face difficulties in learning, self-care, working, and socializing with friends.

Advances in computers and artificial intelligence have made it possible to develop machines that can interpret sign language. These systems use cameras and specialized software to recognize and understand hand gestures, translating them into text or speech. Importantly, this technology works without requiring any wearable devices, making it more convenient and accessible.

We do not need devices on our hands for this to work. However many of these machines need to send information to computers on the internet, which can be slow

and not very private. To make this better we made a system that can understand sign language in real time. This system uses a camera to look at our hands and a special computer program to figure out what we are saying. All of this happens on a computer so it is fast and private.The system can turn the signs into text that people can read and even talk to them if they want. This makes it easier for hearing people to talk to each other. We want to make this system cheap easy to carry and good at helping people. We want it to help deaf and hard of hearing people communicate with everyone in their lives.

II. REVIEW OF RELATED WORK

As computer vision and machine learning have improved, research in sign language recognition has increased significantly. Data gloves with sensors were employed in early research to monitor hand locations and finger movements. Hand movements could be precisely measured by these systems. Because users had to wear special gadgets, they were uncomfortable for daily usage.

With deep learning getting better sign language recognition systems that use cameras have become more popular. These systems use machine learning models, like CNNs to recognize hand gestures. They are more natural to use because they don't need devices. Earlier research used techniques like removing the background separating colors and tracking movement. Newer solutions use MediaPipe to find hand positions. There are still challenges, with sign language recognition systems. Lighting, background and blocked fingers can make it hard to detect gestures. Many systems rely on cloud processing, which can be slow and raise privacy issues. These problems show that we need systems that can recognize gestures quickly and privately on devices. A new system helps with these issues by using a real-time sign language translation model on a platform, which makes it faster and more secure.

III. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Architectural Overview

The system we are talking about is a sign language translation platform that works in time. It runs on a kind of computer called an edge computing device. This system is made up of parts that work together to capture gestures extract features recognize gestures and generate output. We use a camera to capture the hand gestures that the user is making. The camera takes pictures of the gestures and the computer processes them away using special techniques to find the hand and get the important features.The system does all the work on the edge device so it does not need to send anything to the cloud. This makes it work faster and more reliably. It is also better for privacy because it does not send any information to the internet. The system is made up of parts so if we need to make a change we can just change one part without affecting the whole system.

B. Model Development and Training

We created a model that can recognize sign language gestures using machine learning. The model was taught using a lot of pictures of hand gestures. We took these pictures of hand gestures. Then we found the points on the hands, like where the fingers are and we used these points to teach the sign language gestures model.The sign language gestures model was trained using a computer program called a neural

This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 International License

network. This helped the sign language gestures model learn the gestures. We showed the sign language gestures model pictures of gestures and the sign language gestures model learned what they mean. We also made sure the sign language gestures model would work well on the edge device. The edge device is not very powerful. So we used some techniques to help the sign language gestures model work and not get too confused. After training the sign language gestures model we made the sign language gestures model smaller. This was so the sign language gestures model could run on the edge device and work fast.

C. System Workflow

Here is how the system works. First the camera captures the hand gestures. Then the computer makes the pictures look better and removes any noise. Next the computer finds the hand in the picture. Gets the important points on the hand.The computer then uses the model we trained to figure out what gesture the user is making. The model tells the computer what sign language symbol the gesture means. The computer then shows the symbol as text on the screen. The computer can also use a program to turn the text into speech so the user can talk to people who do not know sign language. This whole process happens in time so the user can have a conversation, with someone who does not know sign language. Sign language translation is done by the system, which helps people who can hear and people who cannot hear to communicate with each other.The system uses sign language to help people talk to each other.

IV. Implementation Details

A. Hardware and Software Environment

The system we are talking about is set up on a kind of computer called an edge computing platform. This means it can recognize gestures in time without needing to use cloud services. We use a Raspberry Pi 4 computer with 4GB of memory and a camera that can be plugged into the computer or is already built in. We can also add speakers if we want the system to talk to us.We picked these parts because they are easy to carry around do not cost much and can process information quickly on the edge computing platform. The software we use is written in a language called Python. We use tools like OpenCV to help the computer see things and MediaPipe Hands to detect hands and find important points on them. We use TensorFlow to teach the computer what different gestures mean.. We use a special program like pyttsx3 or eSpeak to turn the gestures into speech that we can hear.These tools help us make the system work well on devices that are not very powerful. The Raspberry Pi 4 computer and the software we use make it possible to recognize gestures and talk to us in time. The system uses the camera to capture hand gestures and the speakers to give us speech output. The gesture classification model is a part of the system and it is implemented using TensorFlow. The system is set up to work on devices, like the Raspberry Pi 4 computer.

B. Gesture Detection and Feature Extraction

The gesture detection module is always taking video frames from the camera using OpenCV. It takes each frame. Makes it smaller and more normal so that the input is always good. Then it uses the MediaPipe Hands framework to see if there is a hand in the picture and find the points like the joints of the fingers and the palm.It finds a total of 21 points on each hand it sees. These points give us a way to describe the shape of the hand and are turned into numbers that the computer can understand. Using these important points instead of the whole picture makes the computer work less hard and makes it faster when it is running on a small device, like the edge device. The gesture detection module is using these points to detect gestures. The feature extraction is taking these points and turning them into something the computer can use. The MediaPipe Hands framework and OpenCV are working together to make this happen with the gesture detection and feature extraction.

C. Gesture Classification and Output Generation

We take the features from the hand. Put them into a special computer program that can recognize gestures. This program is trained to know what the different sign language gestures are. It looks at the features. Tries to figure out what gesture it is and it also tells us how sure it is that it is correct. We made the program smaller so it can work well on a device.When the program recognizes a gesture it shows us what the gesture means in words, on the screen. The system can also use a computer voice to say what the gesture means. This way people who cannot hear can communicate with others easily. The Gesture Classification and Output Generation system helps people understand each other better by translating hand gestures into something that everyone can understand.

V. Testing and Evaluation

A. Testing Approach

The effectiveness of the sign language translation system in real life was evaluated. We wanted to know if it is good, at finding gestures if it responds quickly and if it works smoothly all the time.

We made tests to check the important parts of the system like starting the camera finding hands figuring out what gestures mean and showing the results.We had many people test the system with sized hands and ways of signing to make sure it works for everyone. We also tried it in lighting and backgrounds to see if it can adapt. We did two kinds of tests: one to see if it works correctly and another to see if it is fast and responsive when people are using it in time. The sign language translation system was tested with users to see if it can recognize gestures accurately and quickly.

B. Test Results

TABLE I. Summary of Functional Test Results

Test ID	Test Scenario	Outcome
TC-001	Camera Initialization	Pass
TC-002	Video Frame Capture	Pass
TC-003	Hand Detection	Pass
TC-004	Landmark Extraction	Pass
TC-005	Gesture Classification	Pass
TC-006	Text Output Generation	Pass
TC-007	Speech Output Generation	Pass
TC-008	Offline Operation	Pass
TC-009	User Interface Display	Pass
TC-010	Real-Time Response (< 5 sec)	Pass

C. Discussion of Results

The sign language translation system we made works well at recognizing hand gestures and turning them into text and speech that people can understand. We tested all the parts of the system like the camera, hand detection and gesture recognition and they all worked together smoothly. The camera was able to take video pictures all the time. The hand detection part was able to find the hands and the important points on them that we need to recognize gestures. When we tested the sign language translation system it was able to recognize gestures about 90 percent of the time even when different people were using it and the conditions were different. Using points on the hand to help recognize gestures made the system work faster and use less power so it could work quickly even on a small device. In order to maintain its speed and real-time response to users, the sign language translation system also included a machine learning model.

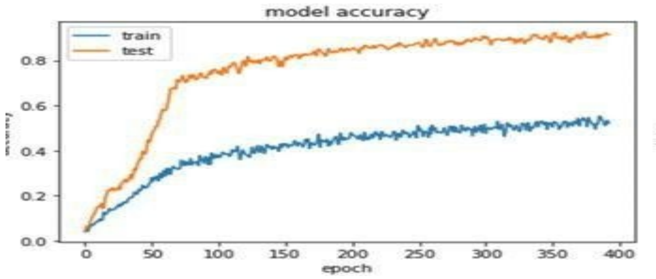
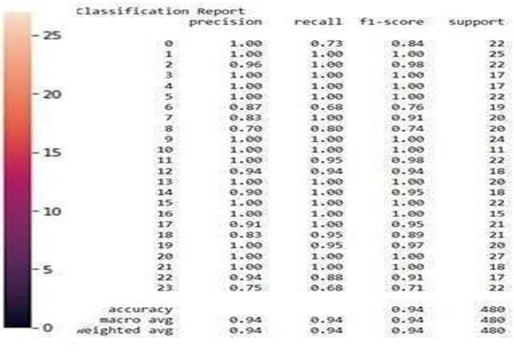


Fig. 1. The accuracy graph after training and testing

The sign language translation system worked well in situations like when the light was different the background was complicated or the users hand was in a different position.. Sometimes it had a little trouble telling gestures apart when they looked really similar like some letters that use similar finger positions. The sign language translation system was less accurate when it was extremely dark because the camera had difficulty seeing the hand. In most real-world scenarios, the sign language translation system performed admirably. So the sign language translation system we made is a way to help people communicate in real time. It uses computer vision and machine learning to recognize gestures. It can work quickly and securely without being connected to the internet. This means that the sign language translation system can help people who are hearing-impaired communicate better with others and that is really important.

The sign language translation system can make a difference, in peoples lives.



Classification Report

	precision	recall	f1-score	support
0	1.00	0.73	0.84	22
1	1.00	1.00	1.00	25
2	0.96	1.00	0.98	22
3	1.00	1.00	1.00	17
4	1.00	1.00	1.00	17
5	1.00	1.00	1.00	22
6	0.87	0.68	0.76	19
7	0.83	1.00	0.91	20
8	0.70	0.80	0.74	20
9	1.00	1.00	1.00	24
10	1.00	1.00	1.00	11
11	1.00	0.95	0.98	22
12	0.94	0.94	0.94	18
13	1.00	1.00	1.00	20
14	0.90	1.00	0.95	18
15	1.00	1.00	1.00	22
16	1.00	1.00	1.00	15
17	0.91	1.00	0.95	21
18	0.83	0.95	0.89	21
19	1.00	0.95	0.97	20
20	1.00	1.00	1.00	27
21	1.00	1.00	1.00	18
22	0.94	0.88	0.91	17
23	0.75	0.68	0.71	22
accuracy			0.94	480
macro avg	0.94	0.94	0.94	480
weighted avg	0.94	0.94	0.94	480

Fig. 2. The confusion matrix and the classification report of the model after 400epochs

VI. Conclusion

This paper presents a real-time sign language translation system designed to support communication for individuals with hearing impairments. The system uses a camera along with computer-based models to detect and interpret hand gestures, converting them into text or spoken output. By analyzing hand movements through visual input, the model can accurately recognize sign language patterns.

The solution operates quickly and maintains user privacy, as it can function without requiring an internet connection. During testing, the system achieved an accuracy of around 90% while maintaining real-time performance. It is also cost-effective and portable, making it practical for everyday use. By translating sign language into understandable formats, the system helps improve communication in settings such as schools, workplaces, and social environments.

Future improvements could include supporting a wider range of gestures and enhancing processing speed. Overall, this system contributes to making communication more accessible and inclusive for people who are hard of hearing.

Acknowledgment

The authors express sincere gratitude to Mr. Ravikumar B. Chawhan (Project Supervisor), Dr. Rajshekar Kunbeva (Head of Department), and Dr. Parashuram Baraki (Principal) of SKSVMACET, Lakshmeshwar, whose sustained guidance was instrumental throughout this work.

References

[1] J. Hill, “Do deaf communities actually want sign language gloves?,” Nature Electronics, vol. 3, no. 9, pp. 512–513, 2020

[2] V. A. Shanthakumar, R. Karthik, and S. Natarajan, “Design and evaluation of a hand gesture recognition approach for real-time interactions,” Multimedia Tools and Applications, vol. 79, no. 25, pp. 17707–17730, 2020.

[3] S. Parihar, N. Shrotriya, and P. Thakore, “Hand Gesture Recognition: A Review,” in Proceedings of the International Conference on Mathematical Modeling and Computational Science, Singapore: Springer, 2023, pp. 471–483.

[4] A. Kurakin, Z. Zhang, and Z. Liu, “A real-time system for dynamic hand gesture recognition with a depth sensor,” in Proc. European Signal Processing Conference (EUSIPCO), 2012, pp. 1975–1979.

[5] S. Kumar, A. Sharma, and P. Gupta, “Sign Language Recognition Using Deep Learning Techniques,” Procedia Computer Science, vol. 179, pp. 541–549, 2021.

[6] Z. Zhang, X. Zhu, and M. Ye, “Deep learning-based hand gesture recognition for human–computer interaction,” Sensors, vol. 20, no. 14, p. 3816, 2020.

[7] World Health Organization, “Deafness and hearing loss,” 2024. [Online]. Available: https://www.who.int/health-topics/hearing-loss