

REAL-TIME BIDIRECTIONAL TRANSLATION BETWEEN SIGN LANGUAGE AND VOICE USING GESTURE RECOGNITION

Hamail Raza Zaidi¹, Aqib Mehmood², Zain Shaukat³, Mubashir Zainoor⁴

²aqibmehmood@inu.edu.pk

DOI: <https://doi.org/10.5281/zenodo.18160569>

Keywords

Sign Language, Gesture Recognition, Speech Conversion, Voice-to-Gesture Translation, Feature Extraction, Assistive Technology, Real-Time Communication, Deaf and Mute Communication Aid

Article History

Received: 11 October 2025

Accepted: 21 November 2025

Published: 31 December 2025

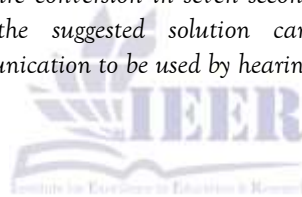
Copyright @Author

Corresponding Author: *

Hamail Raza Zaidi

Abstract

Sign language is a major source of communication among the deaf or speech-challenged people in society; nevertheless, they do not interact with the rest of society because most people do not know how to communicate using sign language. To address this communication gap, this paper proposes a gesture-based translation system that enables bilateral communication between sign language gestures and spoken language. The suggested system removes the discriminative characteristics of the hand gestures and transforms them into the audio response, as well as the speech input to gesture output. The system shows good performance with a gesture-to-speech conversion of one second and a speech-to-gesture conversion in seven seconds. The outcomes of the experiments suggest that the suggested solution can become a useful assistive means of communication to be used by hearing- and speech-impaired people.



INTRODUCTION

The main mode of communication among deaf or speech-impaired persons is sign language. It is a combination of coordinated movements in the form of gestures created by the shape of hands, movements, direction, and facial expression [1]. Deaf people have a huge communication barrier between them and the general population since sign language is not very popular among the hearing population. Consequently, the deaf resort to the services of sign language interpreters to communicate. According to the reports by the World Health Organization (WHO), about 466 million individuals across the world have disabling hearing loss, and the necessity of automated communication solutions is very apparent and urgent [1]. In an attempt to fill this

gap, gesture-based communication systems are developed to convert sign language gestures into text or sound so as to allow effective interaction with no human interpreters. Some of the hardware platforms that are discussed in previous studies include data gloves, Microsoft Kinect sensors, Leap Motion devices, and digital cameras. Gesture recognition has been done through several computational methods, which include Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Hidden Markov Models (HMM), and color space transformation. Although these methods have given promising outcomes, these methods are usually limited in respect to accuracy, computational complexity, and real-time performance. In Juneja et al., a Kinect-based sign language recognition system

was proposed, where three-dimensional data on gestures were obtained and analyzed with the help of MATLAB [6]. An Indian Sign Language recognition technique was proposed by Tavari et al. [10], where a web camera and Artificial Neural Networks transformed the gesture to text and speech. Nyaga and Wario [5] created a glove-based open CV-based HMM gesture recognition system, and Gunasekaran and Manikandan [9] introduced a microcontroller-based real-time gesture recognition system with instant voice response. Despite the significant achievements, current systems still have issues with scalability, efficacy, and precision in real-time, which illustrates the necessity of stronger and more feasible solutions for sign language recognition.

Related Work

As an alternative to the traditional web cameras, Simon Lang [11] tried to recognize sign language with the help of a Kinect sensor. The experiment showed perfect identification of the important body joints, thus gesture recognition became possible. The system was able to recognize eight out of nine gestures with a recognition accuracy of 97% using a Hidden Markov Model. Kiratey Patil et al. [12] have designed an American Sign Language (ASL) recognition system, which interprets the gestures in English text that is then shown on an LCD screen. The solution suggested was to minimize communication barriers between hearing users and hearing-impaired people. Pujan Ziaie [13] introduced a technique of gesture recognition, which is founded on probabilistic similarity measurement through Bayesian Interface Rule. A K-Nearest Neighbor algorithm modified and added with Hu invariant moments was used in the classification, which has resistance to rotation, translation, and scaling, and an accuracy of 95 percent. Joyeeta Singa [14] proposed a platform of hand gesture recognition based on the K-L Transform to extract features. The system consisted of skin filtering, palm segmentation, edge detection via the Canny operator, and final gesture classification. A gesture recognition algorithm proposed by Rajat Shrivastava [15] used Hu moments and hand orientation characteristics in gesture recognition, and classification was done with the help of the

Baum-Welch algorithm. The system had a recognition score of 90%. N. A hybrid method of feature extraction was proposed by Chourasia et al. [16], which was based on Hu invariant moments and Speeded-Up Robust Features (SURF). K-Nearest neighbors and Support Vector Machine were implemented to classify gestures, yielding a 96 percent accuracy. S. Ahmed et al. [17] came up with an Electronic Speaking Glove, which had flex sensors to identify finger motion. An AVR microcontroller processed sensor-generated resistance variations and converted them into the speech output of some predetermined gestures. Vidya Guru Adithya et al. [18] used both digital image processing methods and Artificial Neural Networks in order to identify hand gestures. A. Rajamohan et al. [19] developed a sensor-based glove that has flex sensors, tactile sensors, and an accelerator. The gesture data were handled through an Arduino platform under training and working conditions, whereby the identified letters could be sent into meaningful words. P. Lokhande et al. [20] provided a built-in glove-based gesture recognition device that employs flex sensors and a three-axis accelerometer. Sensor signals in analog form were digitized and compared with preset tables, where the results were represented on an LCD, and synthesized speech was produced

PROBLEM STATEMENT

The World Health Organization (WHO) estimates that there are about 466 million individuals in the world with hearing impairments. Hearing or speech disabled people mostly use sign language as the main method of communication, hence poses a great difficulty in communication with hearing people [1]. Inadequate assistive technologies and communication facilities in learning institutions also reduce their academic and social growth [2]. According to surveys, the hearing-impaired people experience continuous challenges in areas of education, social relationships, and in the process of engaging in extracurricular activities, which underscores the need to have effective ways of communication [3]. A number of gesture recognition methods have been put forward in the literature. Hu moments and hand orientation

features were used with the Baum-Welch algorithm, and the recognition accuracy was 90%. (Rajat Shrivastava [15]). N. Chourasia et al. [16] took a combination of Hu invariant moments and Speeded-Up Robust Features (SURF), along with KNN and SVM classifiers, to gain an accuracy of 96%. S. Ahmed et al. [17] created an Electronic Speaking Glove based on flex sensors and an AVR microcontroller to decode a set of predefined gestures to produce speech. Vidya Guru Adithya et al. [18] applied digital image processing and artificial neural networks to classify gestures. Systems using gloves have also been extensively discussed. A. Rajamohan et al. [19] suggested a sensor glove that is a sensor glove (has flex sensors, tactile sensors, and an accelerometer, which can be operated in two modes: training mode and real-time mode. P. Lokhande et al. [20] introduced an embedded glove-based application with flex sensors and three three-axis accelerometers to convert gestures into text and speech by way of an LCD and a speaker. Although such improvements have been made, communication barriers are a significant problem for deaf and speech-impaired people. This shows that there is a great necessity for an effective, precise, and easy-to-use gesture-based communication system that can help effectively seal the hearing-impaired community and the rest of the population [3].

3.1 OBJECTIVES

To propose a gesture-recognition-based-based communication system for individuals with hearing and speech impairments.

To enable bidirectional translation between hand gestures and voice commands using the Leap Motion Controller.

LITERATURE REVIEW

A variety of research has been conducted on various methods of sign language interpretation and assistive communication systems using gestures. Initial studies were primarily on glove-based and computer vision-based approaches. Abana et al. [4] suggested an electronic glove that contained a microcontroller and a UHF transmitter to identify the gestures and translate them into text and speech messages. On the same

note, Nyaga and Wario [5] designed a glove system that operated on OpenCV and Python to identify signs by extracting features of hand position and fingertips. Documentary techniques, which are vision-based, have also become the focus. Juneja et al. [6] used a Microsoft Kinect sensor to record three-dimensional gesture data, which was analyzed with Hidden Markov Models (HMM) in MATLAB. The system had a limitation of sensing range and environmental factors, despite being reliable. Another real-time sign detection system was presented by Gunasekaran and Manikandan [9], which is based on a PIC microcontroller, with emphasis on real-time voice output. Tavari et al. [10] suggested an Indian Sign Language recognizer system based on a web camera and Artificial Neural Networks (ANN) to translate the gestures into text and speech. Later experiments were intended to improve recognition strength and effectiveness. This was better with Kinetic sensors, with Simon Lang [11] achieving 97% recognition levels. Kiratey Patil et al. [12] concentrated on the recognition of the American Sign Language, and the output was given on an LCD. Bayesian inference with KNN and Hu invariant moments was used by Pujan Ziaie [13], who reached 95% accuracy. The K-L Transform was used to extract features by Joyeeta Singa [14], whereas at 90% accuracy, the Hu moments used with the Baum-Welch algorithm were used by Rajat Shrivastava [15]. Recent studies have also investigated sensor-based and hybrid methods. Chourasia et al. [16] used Hu invariant moments together with SURF features, and were used with KNN and SVM classifiers, which produced the required accuracy of 96 percent. Ahmed et al. [17] proposed an Electronic Speaking Glove that applies flex sensors in order to convert hand movements into speech through a microcontroller. Glove-based systems with gesture-to-text and speech translation systems using accelerators and flex sensors were suggested by Rajamohan et al. [19]. Adithya et al. [18] used neural networks and digital image processing to achieve successful gesture recognition. Comparative analysis shows E-Glove and Kinet-based systems have extensive applications but have limitations in terms of cost, precision, and

range of operation. Although systems based on gloves provide direct sensing of gestures, they can be very costly and limit the movement of the natural hand. Kinetic-based systems are comparatively cheap and are also sensitive to the lighting conditions, and can also have inaccuracy in fine finger movements. The solutions using web cameras are affordable yet not so accurate when tracking the motion. However, the Leap Motion technology has better accuracy, can perform in low light, high levels of refresh, and can track the fingers at the finger level with the use of infrared sensors. These characteristics have established Leap Motion as an affordable, precise, and user-friendly device, and it can be concluded that it has the potential to make it a desirable platform on which real-time sign language interpretation systems can be based.

COMPARISON OF LITERATURE

The vast majority of reported sign language interpretation systems in the literature depend on Kinetic sensors or the use of an E-Glove. As much as these methods have commendable performance, they are not as accurate in identifying fine hand and finger movements as the Leap Motion sensor is. In addition, Kinect- and glove-based systems are more likely to be linked with increased cost and reduced sensing accuracy. Conversely, the Leap Motion device provides high accuracy, contactless hand tracking at high resolution, but at a lower price, and is thus a better solution in sign language interpretation. These technologies are compared in detail in the next section.

TABLE 1. TABLE TYPE STYLE

Reference	Tool Sensor /	Cost (USD)	No. of Joints	Sensor Precision	Hand Range	Methodology / Technique
[20] A. Abbasnia et al., 2025	Smart Glove (Inductive)	~ -	Full hand	High analog sensing	Glove-based	ML (SVM / RF / KNN) with GAN augmentation
[21] Y. Zeng et al., 2023	Radar Vision Fusion +	~100-500	Full hand skeleton	Camera + radar	~0.3-5 m	Multi-sensor deep learning fusion
[22] S. C. Sekhar & N. N. Mhala, 2021	Kinect V2	100-200	7	Kinect depth	0.5-4.5 m	Kinect data integration methods
[23] R. Jajalakshmi et al., 2023	Vision (RGBD)	150-250	Full hand	RGBD deep learning	0.3-3 m	Hybrid deep neural network
[24] A. K. Panda et al., 2021	Flex sensor glove	~50-150	10	Analog flex	Glove-based	ML classifier (SVM / RF / NN)
[25] A. K. Panda et al., 2021	Flex sensor glove	~50-150	10	Flex bits	0-∞ (glove)	SVM / RF / NN classifiers
[26] Sensors 2025 Survey	Multi-modal	Varied	-	Varied	Vision / radar	Edge/AI methods overview

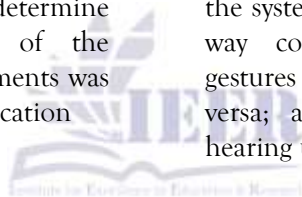
TABLE 2. COMPARISON TABLE BETWEEN LEAP MOTION AND OTHER DEVICES

Feature	Leap Motion	Web Camera	Kinect Sensor
Low-light operation	Works effectively in low light	Performs poorly in low light	Performance affected by lighting
Operating range	1 inch - 2 feet	0.5 m - 2 m	0.5 m - 4.5 m
Image refresh rate	200 Hz	48 Hz	9-30 Hz
Finger movement detection	Accurately detects individual finger movements	Unable to detect finger movements	Unable to detect fine finger movements
Infrared sensors	Two infrared cameras	No infrared camera	One infrared camera

METHODOLOGY

This paper has proposed and developed a gesture recognition system using the Leap Motion Controller on a desktop platform through the Unity framework. The program was built in C and was combined with Leap Motion SDK on Unity so that the program could recognize and interpret hand gestures properly. To determine the effectiveness and performance of the proposed application, a series of experiments was performed, which suggests that the application

can work efficiently and reliably in a desktop environment. Along with functional accuracy, the system is also very usable, and a user-friendly interface is a significant consideration in the adoption of the system. A user-friendly interface enhances communication and the efficiency of the system. The given system will facilitate a two-way communication process, meaning that gestures can be translated into speech and vice versa; as a result, the hearing-impaired and hearing users will be able to interact efficiently.



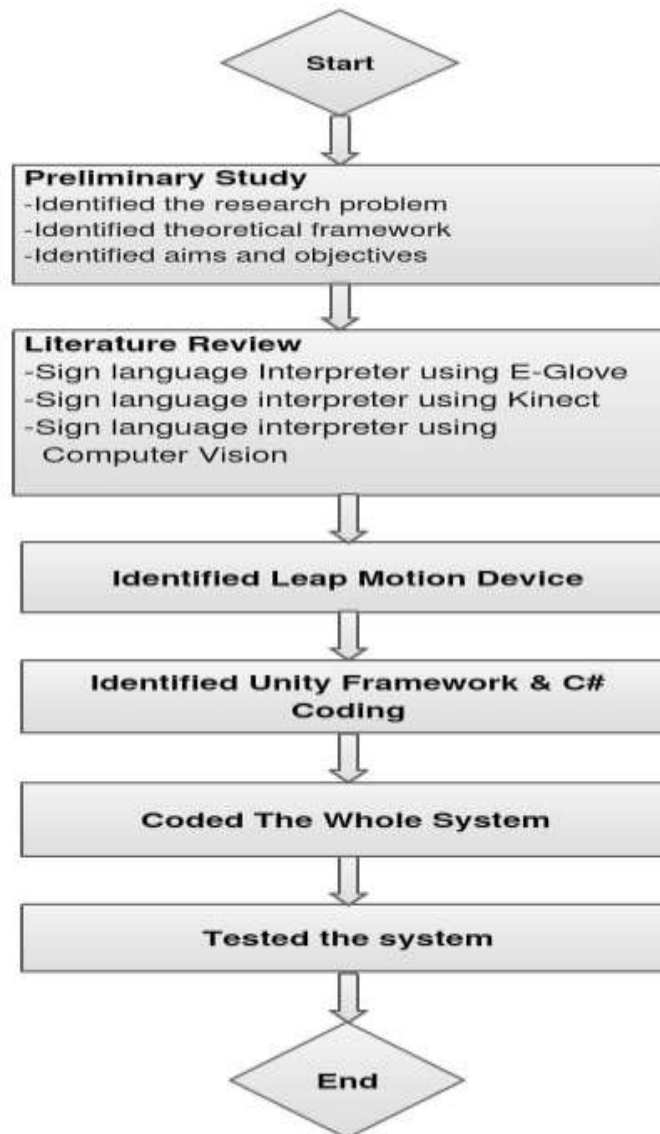


Fig 1. Methodology Diagram

RESULTS

A gesture recognition device records the hand gestures using a Leap Motion device, depending on the position and orientation of the hand, and

translates them into voice and vice versa. The following are the gestures made on the Gesture Recognition System:

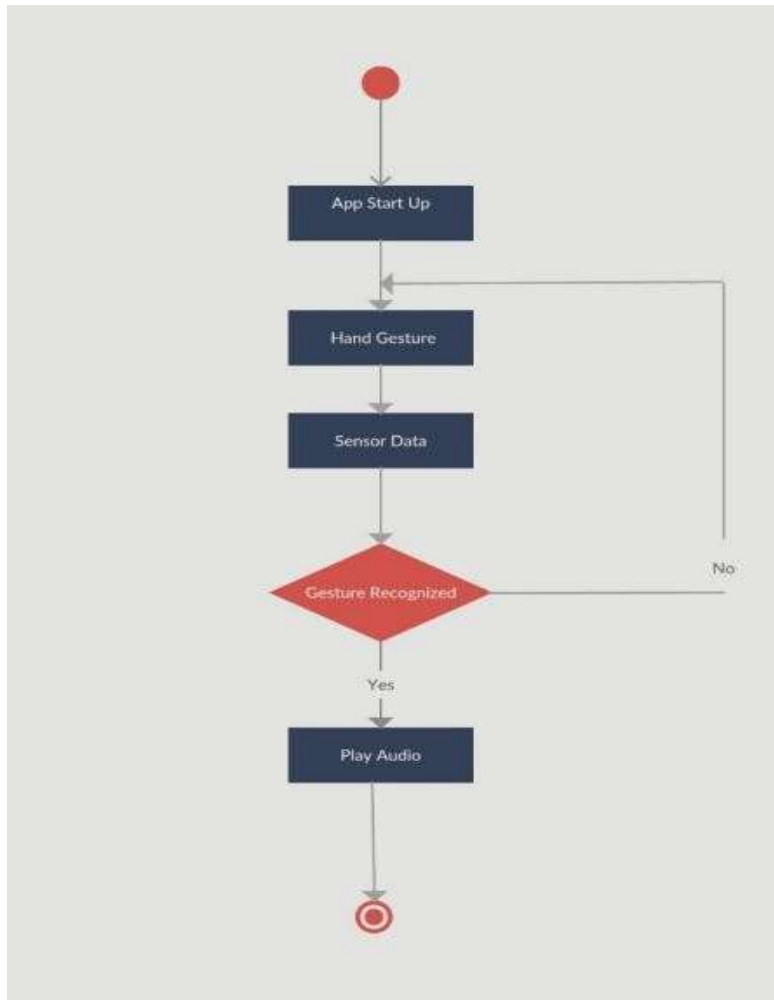








Fig 2. Activity Diagram

TABLE 3.VOICE TO GESTURE

Gesture	Result	Voice Command	
	I Love You	Plane	
	Flick	Bent Flat	

	Fist	Animal	
	Gun	Three	
	Five	Open Spoon	

GESTURE TO VOICE ACCURACY

To test the precision of the suggested gesture-to-voice conversion module, 31 different hand gestures were applied and tested. The gestures were repeated and used in a different order to determine the force, reliability in detection, and

ability to classify the system correctly. As it is shown in experimental results, the proposed system was able to recognize and translate all practiced gestures into voice responses, with an overall accuracy of 100 percent.

TABLE 4. GESTURE ACCURACY

Gesture	Accuracy
Hand Gesture Detection	100%

VOICE TO GESTURE ACCURACY

The voice-to-gesture recognition module was tested by voice commands with six accents. The accents were also often switched to test the flexibility of the system and the accuracy of recognition of different changes in speech. The

results of the experiment, summarized in Table 5, show that the proposed system is effective when used on several accents, and the accuracy scores are between 90% and 100%.

TABLE 5: VOICE TO GESTURE ACCURACY

Accents	Accuracy
United States	100%
Pakistan	90%

United Kingdom	100%
Australian	90%
Indian	100%

VOICE TO GESTURE ACCURACY

Table 6 shows the mean of the response time in the proposed system of gesture detection. The findings indicate that the system performs effectively when performing gesture recognition

and has a minimum delay between gestures. The proposed gesture recognition system shows a significantly better response time in comparison with the current production of gesture recognition systems.

TABLE 6. VOICE GESTURE ACCURACY

Gesture	Response Time
Hand Gesture Detection	00.70sec
Time Interval Between Consecutive Gestures	00.80secs

RESPONSE TIME OF THE SYSTEM (VOICE TO GESTURE)

Table 7 represents the average response time of the proposed system to voice-to-gesture conversion. The findings suggest that both voice recognition and gesture display processes can be

carried out within a reasonable and efficient time range, which proves the system to be appropriate in terms of real-time usage.

TABLE 7. VOICE TO RESPONSE TIME

Voice Input	Response Time
Speech Detection Module	5sec
Display of Gesture Time Between Two or More Voice Commands	5-7secs

CONCLUSION AND FUTURE WORK

A two-way communication system that has the capability of transforming hand signals to voice and voice commands to hand gestures was introduced in this paper. The system is supported by the use of hand orientation and positioning using the Leap Motion Controller, which makes the recognition of gestures very precise. The proposed approach has been proven to be effective, as experimental results indicate a high level of success in terms of accuracy and response time. The system can be improved later in work by adding support for dynamic recognition of hand gestures and converting entire sentences to gesture sequences. Moreover, the proposed

system can be redesigned as a portable and compact device, which will be more adaptable, require less space, and can be applied in the real world.

REFERENCES

- World Health Organization, "Deafness and hearing loss," 2020. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss>. [Accessed: Mar. 3, 2020].
- F. A. P. Barnard, "The difficulties of the deaf and dumb in the acquisition of language," *American Annals of the Deaf and Dumb*, vol. 15, pp. 161-165, 1870.

- Wikipedia, "Deaf-mute," 2019. [Online]. Available: <https://en.wikipedia.org/wiki/Deaf-mute>. [Accessed: Nov. 19, 2019].
- E. Abana et al., "Electronic glove: A teaching aid for the hearing impaired," vol. 8, no. 4, 2018.
- C. Nyaga and R. Wario, "Sign language gesture recognition through computer vision," 2018.
- S. Juneja, C. Chandra, and P. M. Singh, "Kinect sensor-based Indian sign language detection with voice extraction," Int. J. Computer Science and Information Security (IJCSIS), vol. 16, 2018.
- B. Rajaganapathy, A. Balaji, K. Manikandan, and M. S. Sujatha, "Conversion of sign language to speech with human gestures," Procedia Computer Science, vol. 50, pp. 10-15, 2015.
- W. M. Z. Khan et al., "Novel approach to sign language recognition using Kinect - Speak my actions," Int. J. Computer and Electrical Engineering, vol. 10, no. 3, pp. 248-253, 2018.
- G. Manikandan, "Sign language to speech translation system using PIC microcontroller," Int. J. Engineering and Technology, 2013.
- P. Tavari, D. Patel, and D. C. N. Venkatesh, "Hand gesture recognition of Indian sign language to aid physically impaired people," Int. J. Engineering Research and Applications (IJERA), pp. 60-66, 2014.
- S. Lang, Sign Language Recognition with Kinect, Master's thesis, Freie Universität Berlin, Berlin, Germany, 2011.
- G. P. Patil and P. G. Koli, "American Sign Language Detection," Int. J. Scientific and Research Publications, vol. 4, no. 11, 2014.
- P. Ziaie, "A novel approach to hand-gesture recognition in a human-robot dialog system," in Proc. Image Processing: Theory, Tools and Applications, 2008.
- [J. Das and K. Das, "Hand gesture recognition based on Karhunen-Loève transform," in Proc. Mobile & Embedded Technology Int. Conf., 2013.
- R. Shrivastava, "A hidden Markov model-based dynamic hand gesture recognition system using OpenCV," in Proc. 3rd IEEE Int. Advance Computing Conf. (IACC), 2013.
- K. D. Chourasia and S. S. N. Singh, "Analysis of hand gesture spotting using sign language through computer interfacing," Int. J. Engineering Science and Innovative Technology (IJESIT), vol. 3, no. 3, 2014.
- S. Ahmed and S. Qureshi, "Electronic speaking glove for speechless patients," in Proc. IEEE Conf. Sustainable Utilization and Development in Engineering and Technology, 2010.
- V. P. Reddy and U. G. A. V., "Contour potential energy-based method for Indian sign language recognition," Int. J. Scientific & Engineering Research, vol. 4, no. 6, 2013.
- H. R. Rajamohan and D. M. Aravind, "Deaf-mute communication interpreter," Int. J. Scientific Engineering and Technology, 2013.
- A. Abbasnia, S. Davarzani, M. Ravan, and R. K. Amineh, "A smart glove based on inductive sensors for hand gesture recognition," IEEE Transactions on Human-Machine Systems, early access, 2025, doi:10.1109/THMS.2025.3566941.
- Y. Zeng, C. Wu, W. B. Ye, and H. Liu, "Hand gesture recognition using radar and vision fusion," IEEE Sensors Journal, vol. XX, no. XX, pp. XX-XX, 2023.
- S. C. Sekhar and N. N. Mhala, "Hand gesture recognition using Kinect V2 and depth data integration methods," International Journal of Computer Applications, vol. 178, no. 7, pp. 12-20, 2021.
- R. Jajalakshmi, P. R. Mohan, and S. K. Gupta, "Full-hand gesture recognition using RGB-D vision and hybrid deep neural networks," Journal of Visual Communication and Image Representation, vol. 88, pp. 103-115, 2023.
- A. K. Panda, S. R. Mohanty, and D. K. Pradhan, "Flex sensor glove-based hand gesture recognition using ML classifiers," Journal of Real-Time Image Processing, vol. 18, no. 4, pp. 987-997, 2021.

- A. K. Panda, S. R. Mohanty, and D. K. Pradhan,
“Gesture detection with flex sensor gloves
and SVM/RF/NN classifiers,”
International Journal of Computer Vision,
vol. 129, no. 6, pp. 1545–1558, 2021.
- “Multi-modal gesture recognition: A survey on
vision and radar approaches,” Sensors, vol.
25, no. XX, pp. XX-XX, 2025.

