Jurnal J-Innovation Vol.15, No.1, Juni 2026
P-ISSN 2338-2082
E-ISSN 2808-5620

MediaPipe and OpenCV Application for Manipulator Robot Control in Object Retrieval Using Finger-Count Gestures

Rendyansyah[1], Irmawan[2], Caroline[3], Hera Hikmarika[4]
[1,2,3,4] Department of Electrical Engineering, Faculty of Engineering, Universitas Sriwijaya
Jl. Palembang-Prabumulih, Km. 32, Indralaya, Ogan Ilir, Sumatera Selatan
[1]rendyansyah.unsri@gmail.com

ABSTRACT

The development of robotics and computer vision technologies has created opportunities for user-friendly Human-Robot Interaction (HRI) systems. This research aims to design and implement a control system for a 4-degree-of-freedom (4-DOF) manipulator robot that uses finger-count gestures, implemented with MediaPipe and OpenCV. The system uses a webcam to capture real-time hand images, which are then processed by MediaPipe Hands to detect 21 hand landmarks and calculate the number of extended fingers. The detected number of fingers is translated into robot control commands, including moving toward the target, picking up objects, moving objects, releasing objects, and returning to the initial position. The manipulator robot is controlled by an Arduino Uno, which drives the servo motors via serial communication. The test results show that the finger-count gesture detection system achieved an 88% success rate across 50 trials. Meanwhile, the robot testing for picking up and moving objects achieved an 80% success rate across 10 trials. The results of this study demonstrate that integrating MediaPipe and OpenCV can be used effectively as a real-time manual control medium for manipulator robots based on finger-count gestures.

Keywords: *Computer Vision, MediaPipe, OpenCV, Manipulator Robot*

ABSTRAK

Perkembangan teknologi robotika dan computer vision telah membuka peluang dalam pengembangan sistem Human Robot Interaction (HRI) yang mudah digunakan. Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem kendali robot manipulator 4 derajad kebebasan menggunakan isyarat jumlah jari berbasis MediaPipe dan OpenCV. Sistem menggunakan kamera webcam untuk menangkap citra tangan secara real-time, kemudian MediaPipe Hands mendeteksi 21 titik landmark tangan untuk menghitung jumlah jari yang terbuka. Jumlah jari yang terdeteksi diterjemahkan menjadi perintah kendali robot, meliputi bergerak menuju target, mengambil objek, memindahkan objek, melepaskan objek, dan kembali ke posisi awal. Robot manipulator dikendalikan oleh Arduino Uno yang mengatur pergerakan motor servo melalui komunikasi serial. Hasil pengujian menunjukkan bahwa sistem deteksi isyarat jumlah jari mampu mencapai tingkat keberhasilan sebesar 88% dari 50 kali percobaan. Sementara itu, pengujian robot dalam mengambil dan memindahkan objek memperoleh tingkat keberhasilan sebesar 80% dari 10 kali percobaan. Hasil penelitian menunjukkan bahwa integrasi MediaPipe dan OpenCV dapat digunakan secara efektif sebagai media kendali manual robot manipulator berbasis isyarat jumlah jari secara real-time.

Kata kunci: Computer Vision, MediaPipe, OpenCV, Robot Manipulator

DOI : https://doi.org/ 10.55600/jipa.v15i1.379

I. INTRODUCTION

The development of robotics technology in the modern industrial era has grown rapidly, particularly in the fields of automation and human-machine interaction. One of the widely developed technologies is the manipulator robot, a mechanically armed robot designed to perform specific tasks such as moving, picking up, and arranging objects. Manipulator robots have been widely used in the manufacturing, healthcare, education, and research sectors for their ability to enhance work efficiency, precision, and operator safety [1]. Along with technological advancements, manipulator robot systems are no longer exclusively controlled by joysticks or other physical control devices, but have also begun to utilize natural human interaction through body movement and hand gesture recognition [2].

One of the technologies supporting the development of human-robot interaction is computer vision. This technology enables computers to acquire, process, and understand visual information from digital images or videos. Computer vision has been widely used in various fields, including facial recognition, object detection, automated surveillance, and robot navigation systems [3]. In practice, computer vision can serve as a real-time communication medium between humans and robots without requiring additional devices, such as glove sensors or specialized controllers [4]. Using a camera as the primary sensor makes the system more flexible, economical, and easier to implement across various robotics platforms [5].

One widely used method for hand gesture recognition is MediaPipe. This framework, developed by Google, enables real-time hand tracking by detecting landmark points on the hand using a standard RGB camera [6]. MediaPipe Hands offers advantages in processing speed, detection accuracy, and support for mid-range devices [7]. The integration of MediaPipe with OpenCV enables efficient, responsive image processing and hand gesture detection [8]. Previous research indicates that MediaPipe can be utilized for various human-computer interaction applications, such as media player control, computer volume control, presentation navigation, and mobile robot control [9][10].

In the field of robotics, the integration of computer vision with manipulator robots continues to evolve, creating more intuitive Human-Robot Interaction (HRI) systems. Through this approach, human hand movements can be translated into specific commands that directly control the manipulator robot. Such systems allow operators to control the robot without physical contact, thereby enhancing flexibility and user convenience. Furthermore, using hand gestures as a control medium can also minimize dependency on additional hardware that is relatively expensive and complex [12].

Although various studies on gesture recognition and robot control have been conducted, several research gaps remain. Most previous studies only focused on mobile robot control or simple Human-Computer Interaction (HCI) applications, whereas implementations on object-retrieving manipulator robots are still relatively limited [5]. In addition, some studies utilized additional sensors such as glove sensors and depth cameras, which incur higher implementation costs [2]. Other research has also not widely discussed the simple integration of MediaPipe and OpenCV to detect the number of fingers as a control parameter for a manipulator robot in real time, with a high level of responsiveness [8]. The selection of MediaPipe is critical here because its lightweight architecture ensures that the hand-tracking process does not create a computational bottleneck, allowing the robot manipulator to receive and execute commands instantly. Therefore, the novelty of this research lies in the development of an object-retrieving manipulator robot control system based on finger-count detection using real-time MediaPipe and OpenCV integration without additional sensor devices. This system is expected to provide a more natural,

low-cost, and easy-to-implement interaction.

This research aims to design and implement a manipulator robot control system that uses computer vision-based hand gestures, leveraging MediaPipe and OpenCV. Hand gestures, represented by the detected number of fingers, will be translated into specific commands to control the manipulator robot's movements for object retrieval. The benefits of this research are expected to contribute to the development of Human-Robot Interaction (HRI) technology that is more effective, economical, and adaptive to modern automation needs. Furthermore, this research is highly urgent, as it can support the development of smart automation systems in the industrial, educational, and robotics research sectors that require intuitive, minimally physical-contact control methods.

II. RESEARCH METHOD

1. Manipulator Robot

The 4-Degree-of-Freedom (DoF) manipulator robot system in this research is designed to enable hand-gesture-based object retrieval control using computer vision technology. The manipulator robot utilizes four degrees of freedom: rotational movement at the base, forearm movement, upper arm movement, and gripper movement, which functions to clamp objects. The entire system is controlled by an Arduino microcontroller, which drives several servo motors as the primary actuators at each robot joint. The control process is executed in real time via hand gesture detection using a webcam integrated with the MediaPipe framework and the OpenCV library [13].

The 4-DoF manipulator robot is designed using 3D-printed PLA plastic, which is lightweight and strong, making it suitable for retrieving small objects. Figure 1 shows the physical appearance of the 4-DoF manipulator robot used in this study. In the actuator section, servo motors are used to drive each robot joint so that it can move according to the commands from the control system. Each servo motor is connected to an Arduino Uno, which serves as the central controller for all manipulator movements. Additionally, a gripper is attached to the end of the robot arm and is designed to open and close to retrieve specific objects, such as small blocks or other lightweight items [14]. To maintain system stability when all servos move simultaneously, the manipulator robot utilizes an external power supply as the primary voltage source. The hardware schematic of the 4-DoF manipulator robot is shown in Figure 2.

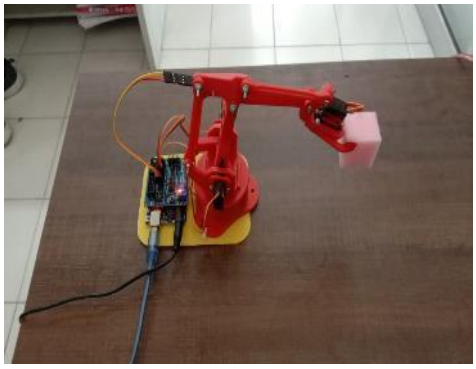


Figure 1. Physical appearance of the 4-DoF manipulator robot.

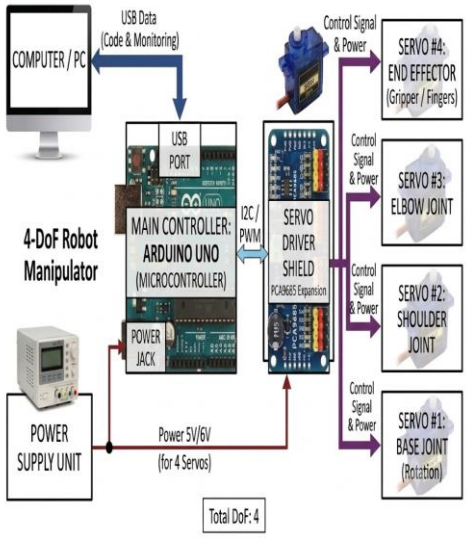


Figure 2. Hardware schematic for the 4-DoF manipulator robot.

2. Robot Control Using Finger-Count Gestures

The software design in this research consists of a computer vision system and a manipulator robot control system that operate in an integrated manner. The computer vision system is developed using OpenCV as the primary library for real-time digital image processing. A webcam is utilized as a visual sensor to capture the user's hand movements, which are subsequently processed by the system. The images captured by the webcam are then analyzed using the MediaPipe Hands framework to detect hand landmarks. MediaPipe can identify 21 hand coordinate points, enabling the system to recognize the position and shape of hand gestures with high accuracy and processing speed [15]. Through this process, the system can calculate the number of extended fingers to issue control commands to the manipulator robot.

The robot control system using finger-count gestures is a computer vision-based Human-Robot Interaction (HRI) technology that allows humans to control robots more naturally, without additional physical devices. In this research, the user's hand movements are captured directly by a webcam, then processed using a combination of MediaPipe and OpenCV to detect hand landmarks and determine the number of raised fingers. The results of gesture detection are subsequently translated into specific commands to control the manipulator robot's movements in real time [16].

The manipulator robot's control mechanism is executed based on the number of fingers detected by the system. One finger is used as a command to move the robot toward the target object, two fingers are used to execute the object-picking process, three fingers are used to move the object toward the destination position, four fingers are used to release the object, and five fingers are used to return the robot to its initial position (original position). The hand gesture detection data is then transmitted from the computer to the Arduino microcontroller via USB serial communication.

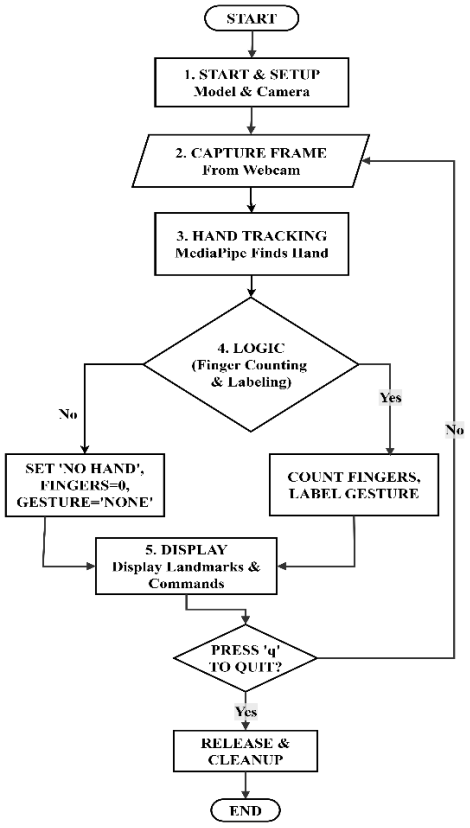


Figure 3. Flowchart of MediaPipe computer vision usage.

The Arduino subsequently processes the data and generates Pulse Width Modulation (PWM) signals to control each servo motor on the manipulator robot according to the received commands. Figure 3 shows the flowchart of using MediaPipe and computer vision for finger-count gesture detection.

The use of finger-count gestures as a control medium offers advantages in creating a more intuitive, flexible, and user-friendly interaction compared to conventional control methods such as joysticks or mechanical buttons. Furthermore, the integration of MediaPipe and OpenCV enables the system to operate in real time with high accuracy using only a standard camera, without requiring additional sensors. This technology holds great potential for application in various fields, such as robotics, industrial automation,

education, and the development of artificial intelligence-based smart control systems.

III. RESULTS AND DISCUSSION

In this research, a 4-Degree-of-Freedom manipulator robot is integrated with a computer to perform automated object picking and moving under manual control via hand gestures. The computer serves as the system's processing hub, running MediaPipe and computer vision programs to detect and recognize the user's hand gestures in real time via a webcam. The detected finger-count gestures are then used as manual control inputs to regulate the robotic arm's movement in response to the user's commands.

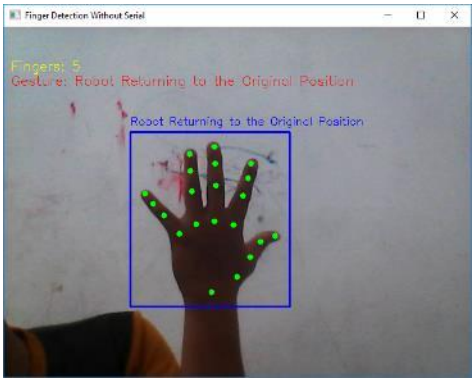


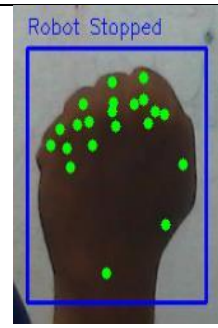
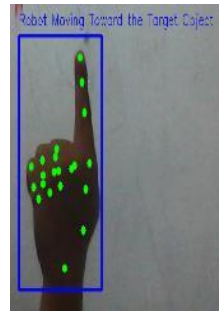
Figure 4. Finger-count gesture detection using MediaPipe and computer vision.

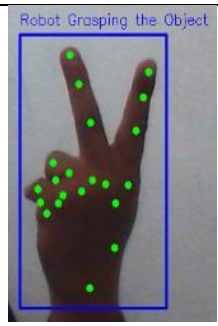
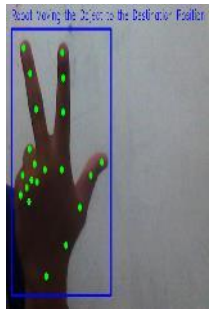
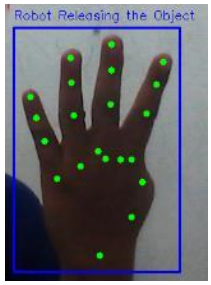
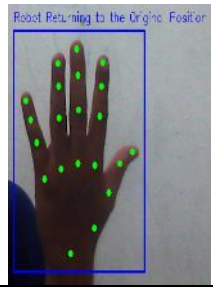
The captured results for finger-count gestures using MediaPipe, as shown in Figure 4, demonstrate that the system successfully detects the user's hand in real time via the webcam. MediaPipe Hands is capable of recognizing the hand's position by displaying 21 landmark points, which are marked with green dots on each critical part of the hand, such as the fingertips, joints, and palm. Additionally, the system displays a blue bounding box to demarcate the hand area detected by the camera.

In the results shown in Figure 4, the system successfully detects five extended fingers, thereby generating the information 'Fingers: 5'. Based on the number of detected fingers, the system translates the finger-count gesture into a command. The system's ability to detect the number of fingers demonstrates that the computer vision method, combining OpenCV and MediaPipe, can effectively perform hand gesture recognition as a control medium for the manipulator robot.

The experiments for detecting the number of fingers from one to five to issue commands for the manipulator robot's movements are shown in Table 1. Finger-count detection is used as a manual control method to regulate the robot's movements according to the user's commands. The system calculates the number of fingers detected by the camera, ranging from one to five fingers, and subsequently translates them into specific instructions for the manipulator robot.

Table 1. Finger-count gesture detection for manual robot commands.

No	Finger-Count Gesture	Command
1		No finger-count gesture. Robot stops.
2		One finger detected. Robot moves toward the target.

3		Two fingers detected. Robot picks up the object.
4		Three fingers detected. Robot moves the object to the destination position.
5		Four fingers detected. Robot releases the object.
6		Five fingers detected. Robot returns to the original position.

All detection results, such as the number of fingers and the robot movement command name, are displayed directly on the monitor in real time. During testing, the distance between the user's hand and the camera is adjusted to ensure the hand position is properly captured, thereby optimizing gesture detection. Additionally, the system uses a 480 × 640 pixel display to present hand detection results and robot command information.

The experiments in this research were conducted for each finger-count gesture used as a movement command for the manipulator robot. Each gesture was tested 10 times to determine the system's success rate in detecting the number of fingers and translating them into robot commands. Based on the test results, the system built with MediaPipe and OpenCV successfully captured the hand and accurately counted the number of fingers in real time.

However, under certain conditions, errors in detecting the number of fingers were still encountered. These errors were generally caused by poor lighting conditions, asymmetrical hand positions, and the hand orientation being read upside down compared to the hand position shown in Figure 4. These factors affected the MediaPipe system's hand landmark reading process, resulting in the detected number of fingers not matching the actual count.

Overall, the integration of MediaPipe and OpenCV in the manipulator robot control system demonstrates reasonably good performance in issuing robot movement commands based on finger-count gestures. Out of 50 trials, the system experienced 6 failures, resulting in an overall success rate of 88%. These results indicate that the MediaPipe- and OpenCV-based computer vision method has strong potential for use in Human-Robot Interaction systems as a real-time control medium for robots.

In this experiment, a MediaPipe- and OpenCV-based computer vision system is used to detect the number of the user's fingers in real time via a webcam. The results of hand gesture detection are translated into specific commands that are transmitted to the manipulator robot to execute actions according to the user's instructions.

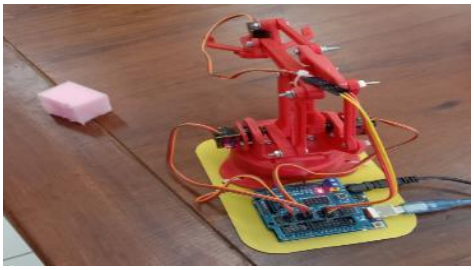
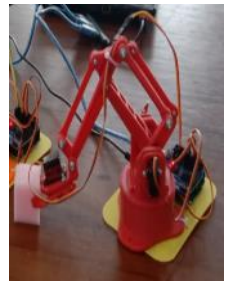


Figure 5. Robotic experimental instruments for picking. and moving objects

The manipulator robot executes each movement based on commands derived from finger-count gestures. Each detected number of fingers represents a specific action for the robot, such as moving toward the object, picking it up, moving it, releasing it, or returning to the initial position. Figure 5 illustrates the robotic experimental setup for picking and moving objects.

In this research, the manipulator robot has five main types of commands controlled using a manual control system based on finger-count gestures. The experiments on the 4-DoF manipulator robot during object picking and movement are shown in Table 2. Testing was conducted to determine the manipulator robot's ability to execute each movement command using the control system that utilizes finger-count gestures. The integration of computer vision technology and the manipulator robot demonstrates that the system can provide real-time robot movement control with a reasonably easy-to-use gesture detection process.

Table 2. Robot experiments for picking and moving objects.

No	Commands	Robot Action
1	One finger detected 	Robot moves toward the target
2	Two fingers detected 	Robot picks up the object
3	Three fingers detected 	Robot moves the object to the destination position
4	Four fingers detected 	Robot releases the object
5	Five fingers detected 	Robot returns to the original position

Based on the testing results shown in Table 2, the experiments of the 4-DoF manipulator robot in picking and moving objects were conducted 10 times. The test results indicate that the robot can execute most commands via the manual control system using finger-count gestures. However, there were two failures during the object-picking process, caused by the gripper's inability to reach the object's position optimally. This condition resulted in the object not being successfully clamped, thereby preventing the moving process from being continued.

Overall, the manipulator robot successfully performed object-picking and moving tasks with an 80% success rate. This success percentage indicates that integrating the MediaPipe- and OpenCV-based computer vision system with the manipulator robot effectively translates finger-count gestures into movement commands. The test results also demonstrate that the robot can pick up and move objects to the destination position via manual control using finger-count gestures, in response to the user's instructions. Consequently, the developed system has the potential to serve as an easy-to-use Human-Robot Interaction method across various robotics applications.

IV. CONCLUSIONS

Based on the research conducted, the integration of MediaPipe and OpenCV was successfully implemented as a real-time manipulator robot control system for finger-count gestures. The system is capable of detecting the number of open fingers using computer vision via a webcam and translating this into robot movement commands: moving toward the target, picking up the object, moving the object, releasing the object, and returning to the initial position. The test results show that the finger-count gesture detection system achieved an 88% success rate across 50 trials. In contrast, the manipulator robot achieved an 80% success rate across 10 trials for object-picking and moving tasks. These results demonstrate that the developed method can support easy-to-use human-robot interaction without requiring additional sensors.

Future research can focus on improving the accuracy of hand gesture detection across various lighting conditions and hand positions. Additionally, implementing artificial intelligence methods and feedback systems can enhance the reliability and precision of robot movements. The integration of computer vision-based object detection can also be developed to enable the robot to recognize and pick up objects more automatically and intelligently.

REFERENCES

[1] V. Moysiadis *et al.*, "An Integrated Real-Time Hand Gesture Recognition Framework for Human–Robot Interaction in Agriculture," *Appl. Sci.*, vol. 12, no. 16, pp. 1–17, 2022, doi: https://doi.org/10.3390/app12168160.

[2] X. Wang, H. Shen, H. Yu, J. Guo, and X. Wei, "Hand and Arm Gesture-based Human-Robot Interaction: A Review," in *ICACS '22: Proceedings of the 6th International Conference on Algorithms, Computing and Systems*, 2022, pp. 1–7. doi: https://doi.org/10.1145/3564982.3564996.

[3] M. Wameed and A. M. Alkamachi, "Hand Gestures Robotic Control Based on Computer Vision," *Int. J. Intell. Syst. Appl. Eng.*, vol. 11, no. 2, pp. 1013–1021, 2023.

[4] S. Togo and H. Ukida, "UAV manipulation by hand gesture recognition," *SICE J. Control. Meas. Syst. Integr.*, vol. 15, no. 2, pp. 145–161, 2022, doi: https://doi.org/10.1080/18824889.2022.2103631.

[5] M. Wameed, A. M. ALKAMACHI, and E. Erçelebi, "Tracked robot control with hand gesture based on mediapipe," *Al-Khwarizmi Eng. J.*, vol. 19, no. 3, pp. 56–71, 2024, doi: https://doi.org/10.22153/kej.2023.04.004.

[6] F. Zhang *et al.*, "MediaPipe Hands: On-device Real-time Hand Tracking," *arXiv Prepr. arXiv2006.10214*, pp. 1–5, 2020.

[7] L. Vaishnavi, A. Kumar, S. Harish, and L. Divya, "MediaPipe to Recognise the Hand Gestures," *WSEAS Trans. Signal Process.*, vol. 18, pp. 134–139, 2022, doi: 10.37394/232014.2022.18.19.

[8] S. N. Budiman, S. Lestanti, S. M. Evvandri, and R. K. Putri, "Pengenalan Gestur Gerakan Jari untuk Mengontrol Volume di Komputer Menggunakan Library Opencv dan MediaPipe," *ANTIVIRUS J. Ilm. Tek. Inform.*, vol. 16, no. 2, pp. 223–232, 2022, doi: https://doi.org/10.35457/antivirus.v16i2.2508.

[9] A. M. Chalik, B. A. Qowy, F. Hanafi, and A. Nuraminah, "Mouse Tracking Tangan dengan Klasifikasi Gestur Menggunakan OpenCV dan Mediapipe," *J. Ilm. Tek. Inform. dan Komun.*, vol. 1, no. 2, pp. 10–18, 2021.

[10] M. Rizaldi, R. Heriansyah, and N. Suhandi, "Pengenalan Gerakan Tangan untuk Kontrol Slide Presentasi Menggunakan Framework Mediapipe, OpenCV, dan Model LSTM," *J. Publ. Tek. Inform.*, vol. 4, no. 3, pp. 98–111, 2025, doi: https://doi.org/10.55606/jupti.v4i3.5332.

[11] N. Kashyap, P. Patharia, and A. K. Kashyap, "Gesture-Based Control of Multimedia Player Using Python and OpenCV," *Int. J. Adv. Technol. Soc. Sci.*, vol. 1, no. 4, pp. 267–278, 2023, doi: https://doi.org/10.59890/ijatss.v1i4.983.

[12] D. Sharma, P. Thakur, S. Gupta, N. K. Dhar, and L. Behera, "Dynamic Hand Gesture Recognition for Robot Manipulator Tasks," in *2024 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2025, pp. 1–6. doi: 10.1109/SMC54092.2024.10831056.

[13] R. N. Fauzan and R. Pramudita, "Controlling a 4 Degree of Freedom (4 DoF) Robot Arm with Hand Gestures Using Computer Vision Technology for Manufacturing Processes," *J. Electr. Electron. Information, Commun. Technol.*, vol. 6, no. 2, pp. 48–54, 2024, doi: https://doi.org/10.20961/jeeict.6.2.90823.

[14] Rendyansyah, B. Y. Suprapto, H. Hikmarika, and Irmawan, "Application of Learning Vector Quantization and Trajectory Planning on a 4-DoF Robotic Arm to Move the Object," *J. Ecotipe (Electronic, Control. Telecommun. Information, Power Eng.*, vol. 10, no. 2, pp. 268–276, 2023.

[15] A. H. Nur'azizan, A. R. Ardiansyah, and R. Fernandis, "Implementasi Deteksi Bahasa Isyarat Tangan Menggunakan OpenCV dan MediaPipe," in *Seminar Nasional Teknologi & Sains*, 2024, pp. 331–337.

[16] A. Y. Prasetya, Subairi, and A. Rabi, "Deteksi Gestur Tangan Menggunakan Computer Vision Untuk Pengendalian Robot Lengan," *J. Mech. Electr. Technol.*, vol. 5, no. 1, pp. 18–25, 2026.