JOURNAL OF ARTIFICIAL INTELLIGENCE AND ROBOTICS
Apr-Jun 2025, VOL. 2, ISSUE 2, pp. 11-24
https://doi.org/10.61577/jaiar.2025.100007

RESEAPRO JOURNALS

ORIGINAL ARTICLE

OPEN ACCESS

AI-Vision empowered gesture recognition with 21 hand landmarks for Xtensa-powered IoT autonomous robot navigation

Akash V. Katode[1] and Neha Dagar[2]

[1]Deputy Manager/ STEM Trainer – STEM Education, Myra Academy, Hyderabad, Telangana, India
[2]National Head – STEM Education, Myra Academy, Hyderabad, Telangana, India

ABSTRACT

The integration of Artificial Intelligence (AI) and the Internet of Things (IoT) is reshaping robotics into intelligent, adaptive, and human–interactive systems. In this work, we, as part of the Research and Development team at Myra Academy, present an IoT and AI-vision enabled gesture recognition framework for intelligent robot navigation. Dedicated to advancing STEAM education, Myra Academy has developed this system as a superior alternative to traditional methods such as Standard Firmata, which connects Python environments to embedded systems via serial communication, and Bluetooth-based approaches for AI–embedded interfaces. These older techniques are limited in range, speed, and reliability, whereas our design exploits IoT-enabled Wi-Fi connectivity powered by the Xtensa LX6/LX7 processor architecture, offering robust long-range communication and improved accuracy through AI-vision strategies. The proposed framework employs computer vision techniques with OpenCV and Mediapipe to detect and track hand landmarks, while an AI-driven decision module interprets gestures to control the robotic platform in real time. Experimental evaluation confirms low-latency transmission, reliable gesture detection under varying lighting conditions, and smooth navigation performance. Beyond its technical capabilities, this innovation serves as a powerful educational tool, enabling students to explore IoT and AI-based robot navigation with enhanced connectivity and intuitive interaction. By bridging advanced AI-vision techniques with IoT frameworks, the system has the potential to redefine STEAM education, strengthening AI and robotics curriculum while promoting innovation in smart environments and assistive robotics.

KEY WORDS

Open CV, MediaPipe; AI Robot; Robotic Navigation, Computer vision; AI Robotic Navigation; IOT Robot

ARTICLE HISTORY

Received 21 April 2025;
Revised 12 May 2025;
Accepted 21 May 2025

Introduction

Traditional robotic navigation techniques predominantly rely on predefined instructions, fixed sensor thresholds, or rule-based control algorithms, such as those implemented via microcontroller-based systems with accelerometers, gyroscopes, or Standard Firmata interfacing through serial communication. While functionally adequate for simple tasks, these approaches are inherently constrained by their limited adaptability, low environmental awareness, and dependence on complex hardware integration. Conventional sensor-driven systems often require multiple discrete modules, intricate wiring, and meticulous calibration, imposing substantial challenges for implementation, particularly in educational contexts where intuitive interaction is critical. Furthermore, these methods exhibit reduced precision and robustness, as they are susceptible to sensor noise, latency, and limited resolution, while communication protocols such as serial or Bluetooth restrict data throughput and operational range, limiting real-time responsiveness and multi-device scalability. In contrast, AI-driven navigation frameworks leverage advanced computer vision, machine learning, and sensor fusion techniques, enabling robots to perceive and interpret dynamic environments intelligently, adapt decision-making in real time, and maintain high operational accuracy with minimal hardware complexity. By integrating AI with IoT-enabled connectivity, such systems achieve seamless remote control, long-range communication, and scalable interaction, effectively bridging the gap between sophisticated robotics and user-friendly, educationally accessible platforms. Consequently, AI-based navigation not only enhances robustness, adaptability, and precision but also facilitates intuitive human robot interaction, positioning it as a superior paradigm for contemporary robotics education, research, and practical deployment [1-3].

Artificial Intelligence (AI) has emerged as a transformative technology across diverse domains, revolutionizing processes through its capability to analyze complex data, learn from patterns, and enable autonomous decision-making. In robotics, AI facilitates intelligent perception, adaptive control, and real-time interaction, which are essential for creating systems capable of performing complex tasks with minimal human intervention. The significance of AI is particularly pronounced in robotics navigation, where conventional control mechanisms lack adaptability, and precise real-time decision-making is crucial for accurate motion and obstacle avoidance. AI-driven navigation systems leverage advanced computer vision, machine learning, and sensor fusion to interpret the environment and respond dynamically, enabling more intuitive and reliable human robot interaction [2, 3].

Parallelly, the Internet of Things (IoT) has fundamentally reshaped connectivity paradigms, allowing embedded devices and sensors to communicate seamlessly over networks. IoT-enabled robotics systems can transmit and receive commands in real time, access cloud-based data,

***Correspondence:** Mr. Akash V. Katode, Department of Biomedical Engineering, National Institute of Technology Raipur, Chhattisgarh, India,
e-mail: Kstodeakash@gmail.com

and facilitate remote monitoring and control. The integration of AI and IoT creates cyber-physical systems where intelligent decision-making and ubiquitous connectivity converge, offering a foundation for autonomous, responsive, and scalable robotic platforms. Such systems are not only technologically superior but also essential for educational applications, where interactive, hands-on experiences are crucial for learning [4].

The need for AI-enabled navigation in robotics is particularly relevant in the context of human–robot interaction. Conventional robotic control techniques, such as Standard Firmata, rely on serial communication to interface Python environments with embedded systems, requiring complex understanding and hardware setup. Similarly, Bluetooth-based approaches offer limited transmission range, lower data throughput, and inconsistent reliability. Traditional microcontroller (MCU) and ADXL sensor-based gesture-controlled systems pose additional challenges due to hardware complexity, sensor calibration issues, and difficulty of implementation for young learners, particularly in K–12 environments where intuitive interaction is critical. These limitations restrict the practical adoption of robotics as an educational tool, emphasizing the necessity for simpler, reliable, and engaging interfaces. In parallel, STEAM education encompassing Science, Technology, Engineering, Arts, and Mathematics has emerged as a pivotal approach to shaping 21st-century learning. Globally, STEAM initiatives are redefining education by integrating hands-on, interdisciplinary learning experiences, fostering creativity, critical thinking, and problem-solving skills among students from early childhood through K-12. Robotics plays a central role in STEAM education by offering interactive, tangible learning experiences, yet its potential is often limited by traditional control methods that are difficult for students to implement and understand.

Recognizing these challenges, the Research and Development team at Myra Academy has designed an innovative IoT and AI-vision enabled gesture recognition framework for intelligent robot navigation. Figure 1 illustrates the design of our innovative approach and its strategic operational methodology, highlighting the key functional components and their interconnections within the system.

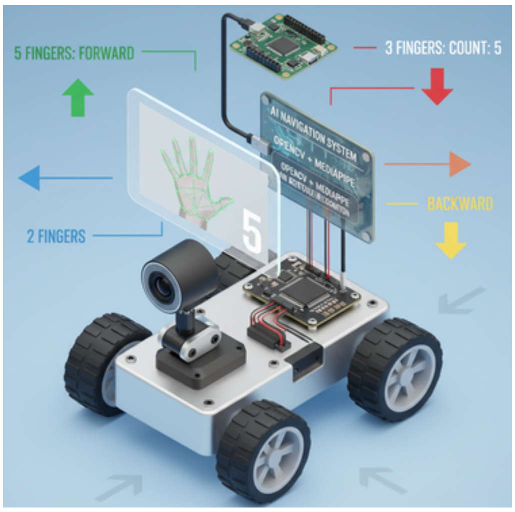

Figure 1. Design of our innovative approach and its strategic operational methodology.

Leveraging the Xtensa LX6/LX7 processor architecture, the system integrates AI-based computer vision techniques with IoT-enabled Wi-Fi connectivity, providing robust long-range communication, low latency, and high-accuracy gesture recognition. This framework not only overcomes the drawbacks of traditional methods such as the complexity of Firmata, limited Bluetooth range, and cumbersome hardware but also offers a scalable, intuitive, and engaging platform for STEAM education, allowing students to explore AI and IoT concepts effectively. Through this research, we aim to demonstrate that combining AI-vision strategies with IoT infrastructure can redefine robotics education, enhance practical learning in STEAM curriculum, and provide a technically robust yet accessible platform for hands-on exploration of autonomous robot navigation. The proposed system represents a significant step toward bridging advanced robotics technology with educational innovation, fostering the next generation of learners and innovators.

Objectives

To achieve the overarching goal of developing an innovative and effective AI-vision and IoT-enabled robotic navigation system, it is essential to define clear research objectives. These objectives serve as guiding principles for the design, implementation, and evaluation of the proposed framework, ensuring that both technical performance and educational impact are systematically addressed. The key objectives of this research are outlined as follows:

1. To develop an AI-vision enabled gesture recognition framework that can accurately detect and interpret human hand gestures in real time for robotic navigation.
2. To integrate IoT-enabled wireless communication using the Xtensa LX6/LX7 processor architecture, ensuring robust, low-latency, and long-range connectivity between the gesture recognition system and the robotic platform.
3. To address limitations of traditional robotic control methods, such as Standard Firmata, serial communication, Bluetooth-based interfaces, and sensor-based navigation systems, by providing a simplified, accurate, and scalable alternative.
4. To create an intuitive and interactive platform for STEAM education, enabling students from K–12 and beyond to explore AI, robotics, and IoT concepts through hands-on, contactless learning experiences.
5. To evaluate the performance of the proposed system, including gesture recognition accuracy, response latency, reliability under varying environmental conditions, and effectiveness as an educational tool.
6. To demonstrate the potential of AI–IoT convergence in enhancing human–robot interaction, making robotic navigation more accessible, adaptable, and educationally relevant.

Literature Review

In recent years, advancements in robotics, AI, and computer vision have transformed how autonomous robots perceive, navigate, and interact with their environments. Researchers have gathered data from experimental testbeds, simulation platforms like Gazebo, PyBullet, and Habitat-Sim, and datasets such as the Cornell Grasping Dataset, Yale-CMU-Berkeley Object Dataset, and Multi-Illumination dataset. These resources provide standardized benchmarks for

https://doi.org/10.61577/jaiar.2025.100007

designing and validating algorithms for localization, mapping, gesture recognition, and navigation, while middleware frameworks like ROS offer modular environments for testing robotic control strategies.

The shift toward vision-based techniques enables more flexible and cost-effective solutions, allowing robots to interpret gestures, recognize objects, and construct detailed maps. Methods like MediaPipe landmark detection, CNNs, RNNs, reinforcement learning, and feature-based pose estimation have been widely explored to enhance navigation and human–robot interaction [5, 6].

This section reviews contributions in robotic grasping in space, industrial AGVs, gesture-controlled robotic arms, Turtlebot SLAM navigation, and RGB-D indoor navigation [7], highlighting challenges and solutions that advance vision-driven robotic autonomy across domains from manufacturing to assistive technologies.

Human-Machine Interaction: A Vision-Based Approach for Controlling a Robotic Hand Through Human Hand Movements

Research by García-Gil et al., titled "Human-Machine Interaction: A Vision-Based Approach for Controlling a Robotic Hand Through Human Hand Movements," presents an advanced gesture-controlled robotic hand that operates without traditional sensors or physical interfaces. Using MediaPipe and computer vision, the system interprets human hand movements and translates them into control commands for a microcontroller. The robotic hand, equipped with six servomotors (five for fingers and one for the wrist), features an orthonormal configuration allowing 180° wrist rotation without occlusion. Relying on a simple 2D camera, it achieves up to 95% accuracy in gesture replication, offering a cost-effective and intuitive solution for human–robot interaction in assistive robotics and teleoperation [8].

Enhanced Route navigation control system for TurtleBot using human-assisted mobility and 3-D SLAM optimization

Research by Kumar et al., titled "Enhanced Route Navigation Control System for Turtlebot Using Human-Assisted Mobility and 3-D SLAM Optimization," presents an intelligent, power-assisted Turtlebot designed to aid human mobility. The robot moves between predefined positions based on user instructions, remaining stationary in the absence of input for safety. A rotating Kinect sensor provides environmental perception, while 3-D SLAM with a Kalman filter enhances localization and mapping accuracy. Experiments on a U-shaped path showed reliable forward and reverse navigation, with reverse motion achieving 5–6% better performance. The study demonstrates the potential of such dynamic control systems for future wheelchair navigation and mobility assistance applications [9].

Vision-based omnidirectional indoor robots for autonomous navigation and localization in manufacturing industry

Research by Patruno et al., titled "Vision-Based Omnidirectional Indoor Robots for Autonomous Navigation and Localization in Manufacturing Industry," presents the development of omnidirectional automated guided vehicles (omniAGVs) for intelligent material transport in industrial settings. The system integrates vision-based perception with cognitive robotics to enable autonomous navigation and human interaction without major facility changes. Key modules include combined visual and wheel odometry for localization, obstacle detection and classification for adaptive navigation, and tag detection for cart-picking and global positioning. Experimental results confirm reliable and efficient autonomous navigation, highlighting the role of vision-based perception in advancing smart manufacturing robotics [10].

Machine learning in robotic grasping control in space application

Research by Jahanshahi and Zhu, titled "Review of Machine Learning in Robotic Grasping Control in Space Application," explores the integration of machine learning techniques in robotic grasping for space environments. The study highlights how methods like deep learning, reinforcement learning, CNNs, RNNs, and transfer learning have advanced robotic perception and adaptability beyond conventional grippers. It discusses challenges such as microgravity, occlusions, and the sim-to-real gap, emphasizing how ML enhances robustness and multimodal data fusion. The authors also note the importance of datasets (e.g., Cornell, Yale-CMU-Berkeley) and simulators (Gazebo, PyBullet) for training and testing. The paper concludes with future directions in multimodal learning, sim-to-real transfer, and swarm robotics for autonomous space manipulation [11].

Hand Gestures Replicating Robot Arm based on MediaPipe

Research by Altayeb, titled "Hand Gestures Replicating Robot Arm Based on MediaPipe," presents a vision-based robotic arm that replicates human hand movements through gesture recognition. Using Python and MediaPipe for detection and an Arduino Nano to control five servo motors, the system achieves real-time gesture replication with up to 96% accuracy. This study demonstrates an intuitive, accurate, and cost-effective approach for human–machine interaction and robotic control [12].

Efficient learning-based robotic navigation using feature-based rgb-d pose estimation and topological maps

Research by Rodríguez-Martínez et al., titled "Efficient Learning-Based Robotic Navigation Using Feature-Based RGB-D Pose Estimation and Topological Maps," introduces a cost-effective and robust indoor navigation framework. The system combines feature-based pose estimation with an MLP policy, using RGB-D keyframes from human-driven traversals to build topological maps based on visual similarity and motion constraints. During operation, LightGlue and SVD estimate six-DoF poses, while real-time graph editing and Dijkstra replanning manage obstacles or low-visibility scenarios. Experiments in multiple simulated environments showed precise navigation within 0.1 m of targets, confirming the method's efficiency and adaptability without relying on metric SLAM or extensive training data [13].

https://doi.org/10.61577/jaiar.2025.100007

RESEAPRO JOURNALS

Agency bands with measurable thresholds

Table 1. Summary table of research gaps.

Research Area	Current Focus	Identified Gaps	Future Opportunities
Human–Robot Interaction	Vision-based hand gesture (Media Pipe, 2D camera)	Lacks multimodal input, adaptive learning	Integrate voice/EMG/tactile + RL-based adaptation
Navigation (SLAM & Pose Estimation)	3D SLAM, RGB-D features	No semantic/contextual awareness	Add semantic mapping & human-intent prediction
Robotic Grasping (Space)	ML-based grasp planning	No cross-domain validation or real-time adaptability	Transfer learning + on-orbit adaptation
Industrial AGVs	Vision-based localization	Limited scalability, lacks energy optimization	Multi-robot coordination + energy-aware algorithms
Overall Evaluation	Lab-based validation	No standard datasets or benchmarks	Develop unified benchmarking framework

Methodology

Basic blocks of system design

The proposed gesture-controlled robotic system is systematically structured into three principal stages: Input and Vision Processing, Wireless Communication with IoT, and Robotic Control and Navigation. Each stage comprises multiple functional sub-blocks, designed to ensure accurate gesture recognition, efficient wireless transmission, and precise robotic actuation. Figure 2 illustrates the basic and important three blocks of the methodology block architecture, providing a visual representation of the system's core functional components and their interconnections.

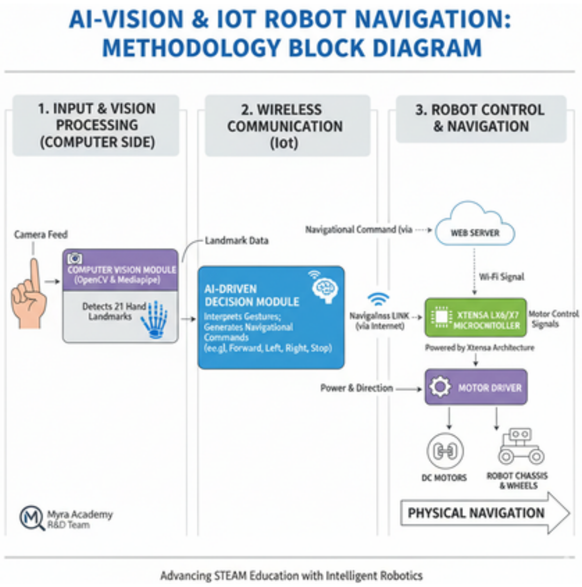

Figure 2. Basic and important three blocks of the methodology block architecture.

The detailed block-level description of the methodology is as follows:

Input and vision processing

The initial stage of the system is dedicated to capturing and interpreting user input through advanced computer vision techniques. A high-resolution camera continuously streams real-time video of the user's hand, providing the raw visual data for subsequent processing. This video feed is analyzed using a Computer Vision Module powered by frameworks such as OpenCV and Mediapipe, which performs noise reduction , background segmentation, and hand region extraction. The module detects and tracks 21 distinct anatomical landmarks on the hand, including fingertips, joints, and the palm, providing accurate spatial coordinates and motion trajectories. These landmark data are then interpreted by an AI-Driven Decision Module, which employs trained machine learning algorithms to classify both static and dynamic hand gestures. By analyzing geometric configurations and temporal motion patterns, this module translates complex hand poses into discrete, unambiguous navigational commands such as "move forward," "turn left," or "stop." This stage ensures high accuracy, robustness to varying lighting conditions, and reliable gesture recognition for effective human-robot interaction.

Wireless communication with IoT

The second stage bridges the computer-based gesture recognition system and the robotic platform via a robust IoT-enabled wireless communication link. Commands generated by the AI-Driven Decision Module are transmitted to a Web Server, which functions as a central communication hub. The server encodes, organizes, and relays the command data over Wi-Fi, providing low-latency, reliable transmission with greater range and speed compared to conventional serial or Bluetooth links. The transmitted commands are received by the Xtensa LX6/LX7 Microcontroller, which decodes the incoming data and prepares it for execution by the robotic actuators. This wireless communication framework ensures seamless real-time interaction, allowing gesture inputs to be converted into instantaneous robot responses while maintaining system stability and minimizing transmission delays.

Robotic control and navigation

The final stage involves converting the received commands into precise physical movement of the robot. The Xtensa LX6/LX7 Microcontroller functions as the Command Processing Unit, generating appropriate low-level control signals based on the interpreted commands. These signals are sent to the Motor Driver Interface, which acts as a power conversion stage, delivering the required voltage and current to the DC Motors mounted on the robot chassis. The motors, in turn, rotate the wheels, propelling the robot and enabling smooth directional changes and speed regulation. The Robot Chassis provides structural support for all components, ensuring stability, weight distribution, and alignment for accurate motion translation. Optionally, a feedback module can monitor wheel rotation, orientation, and position, enabling closed-loop control for enhanced precision, obstacle avoidance, and safety. This stage guarantees a direct and proportional response, translating

https://doi.org/10.61577/jaiar.2025.100007

human hand gestures into reliable and controlled robotic navigation.

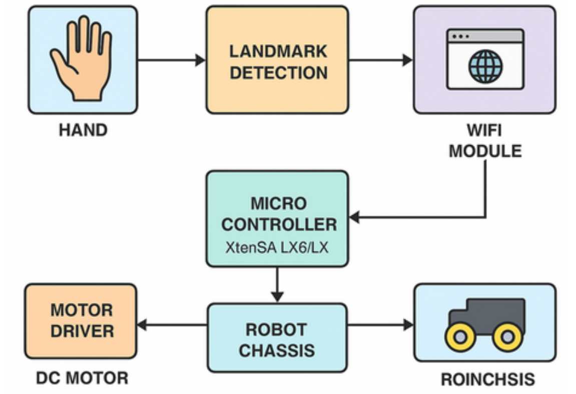

Figure 3. Block diagram of system.

Detail subblocks of methodology

The proposed system consists of multiple functional blocks, each contributing to the seamless operation of the gesture-controlled robotic platform. The methodology is primarily structured around the following blocks. The proposed system is further divided into multiple sub-blocks as part of the methodology, as illustrated in Figure 3, which provides a detailed view of the internal functional components and their interactions.

Detailed block-level description of methodology

Gesture input and hand landmark detection

The gesture input block serves as the primary interface for human-robot interaction and constitutes the foundational element of the system. In this stage, hand gestures are captured using a high-resolution camera integrated into a computer vision system. The captured frames are analyzed using advanced landmark detection algorithms capable of identifying 21 distinct anatomical key points on the hand, including fingertips, joints, and the palm. These key points provide precise spatial coordinates and dynamic trajectories, allowing accurate modeling of finger articulation, hand orientation, and overall motion patterns. The landmark detection framework leverages state-of-the-art techniques to ensure robustness against occlusions, variable lighting conditions, and diverse hand orientations. By providing reliable, high-precision hand tracking, this block establishes a critical foundation for subsequent processing, enabling the system to interpret complex static and dynamic gestures in real time with minimal latency and high accuracy [14-16].

AI-Vision processing

Following the extraction of hand landmarks, the AI-Vision Processing block performs sophisticated analysis to convert raw landmark data into actionable commands. This block incorporates machine learning and deep learning models for feature extraction, pattern recognition, and gesture classification. Temporal analysis techniques are applied to ensure smooth interpretation of dynamic gestures and to minimize noise and erratic fluctuations in the input data. The module evaluates geometric configurations of the hand, including inter-finger angles, joint trajectories, and relative landmark distances, translating these into discrete control instructions such as “move forward,” “turn left,” or “stop.” By implementing advanced filtering algorithms and gesture smoothing techniques, the AI-Vision Processing block bridges the gap between human intent and robotic actuation, generating structured, reliable digital commands while preserving responsiveness and accuracy [15].

IoT communication / web server interface

The IoT Communication block provides a high-speed, reliable, and secure wireless communication channel between the gesture recognition system and the robotic platform. The processed commands from the AI-Vision Processing block are transmitted to a dedicated Web Server, which functions as the central communication hub. The server handles command encoding, packetization, transmission, and acknowledgment, ensuring low-latency delivery even over variable network conditions. Utilizing Wi-Fi-based IoT protocols, this block offers superior range, reliability, and throughput compared to traditional Bluetooth or wired serial communication. The Web Server ensures that data packets are accurately relayed to the robot-mounted Xtensa LX6/LX7 Microcontroller, maintaining synchronization between gesture input and robot actuation [17]. This real-time transmission infrastructure is crucial for responsive system performance, particularly in applications requiring instantaneous reaction to user gestures.

Xtensa LX6/LX7 microcontroller

The Xtensa LX6/LX7 microcontroller acts as the central computational unit of the robotic system, responsible for interpreting incoming commands and orchestrating coordinated actuation. It executes low-level control logic, manages timing sequences, and interfaces with multiple peripherals including motor drivers, sensors, and optional feedback modules. Leveraging its high-speed processing capability and real-time operational efficiency, the microcontroller converts digital commands into precise electrical signals for motor control, ensuring robust and responsive robot navigation [18]. Its integration enables modular scalability, allowing additional sensors or actuators to be incorporated without significant redesign.

Motor driver and dc motor interface

The motor driver block serves as an essential intermediary between the microcontroller and the DC motors. It converts low-power digital signals from the microcontroller into high-power analog signals, providing controlled voltage and current to drive the motors [19]. This block allows bidirectional motor control, speed modulation, and torque adjustment, enabling the robot to execute precise maneuvers and maintain stable motion over uneven surfaces. By integrating advanced pulse-width modulation (PWM) techniques and protective circuitry, the motor driver ensures operational safety, power efficiency, and smooth actuation under dynamic load conditions.

Robot chassis and wheels

The chassis and wheel assembly constitutes the structural framework and mobility system of the robot. The chassis is engineered for mechanical robustness, balance, and precise alignment of motors, sensors, and communication modules. It provides the physical platform for all electronic components while ensuring weight distribution that supports stable navigation. The wheel configuration, in conjunction with the motor driver and DC motors, enables differential or omnidirectional movement, facilitating accurate trajectory tracking, turning, and obstacle negotiation. Structural design considerations include vibration damping, low friction, and

https://doi.org/10.61577/jaiar.2025.100007

ease of maintenance, ensuring long-term reliability of the robotic platform [20].

Feedback and monitoring (Optional)

Although not explicitly required for basic operation, an optional feedback and monitoring subsystem can be integrated to enhance precision and safety. Sensors mounted on the chassis can track parameters such as wheel rotation, orientation, and position, providing real-time feedback to the microcontroller. This closed-loop control allows for adaptive motion correction, improved navigation accuracy, and collision avoidance [21]. Additionally, feedback data can be transmitted back to the IoT server for monitoring or analytics, enabling system diagnostics, performance optimization, and remote supervision.

Implementation

In this section, we delineate the detailed implementation of the proposed gesture-controlled robotic system, encompassing both software and hardware development, to realize a functional prototype. The implementation integrates advanced computer vision techniques, AI-driven gesture recognition, wireless communication, and embedded control to achieve seamless robot navigation based on human hand gestures [18-21].

The software implementation begins with capturing real-time visual data using a standard computer webcam. The video stream serves as the primary input to the computer vision system, which is implemented using the OpenCV framework in conjunction with MediaPipe, a state-of-the-art machine learning library specialized for hand tracking and landmark detection [16,19]. MediaPipe enables the precise identification of 21 anatomical hand landmarks, including fingertips, joints, and the palm center, providing both spatial and temporal information about hand configurations. These landmarks constitute the fundamental parameters for gesture interpretation, allowing the system to distinguish subtle variations in hand posture and motion dynamics [22].

Using the landmark coordinates, a custom AI-driven decision-making module is implemented to classify gestures into discrete control commands for robot navigation. The module is developed in a Python programming environment, where feature extraction, normalization, and geometric analysis of inter-landmark distances are performed to differentiate between gestures such as “move forward,” “turn left,” “turn right,” and “stop.” Each gesture is mapped to a corresponding digital command, which is then transmitted over the internet via a dedicated web server. The server functions as a central hub, ensuring low-latency, reliable, and secure delivery of the command signals to the robotic platform [14, 22].

On the hardware front, the robotic system is designed around the Xtensa LX6/LX7 microcontroller, which serves as the core embedded controller. The microcontroller firmware is developed to interpret incoming commands from the web server and generate corresponding control signals for the motor driver circuit [23]. The L298N motor driver is interfaced with the DC motors to facilitate bidirectional rotation, enabling precise movement of the robot chassis. The embedded firmware incorporates real-time motor control algorithms , pulse-width modulation (PWM) for speed regulation, and direction control logic to execute clockwise and anticlockwise rotation based on the received commands. The firmware is burned onto the Xtensa microcontroller to integrate all functional modules, ensuring the coordinated operation of motors, sensors, and communication interfaces [24].

The complete system design, including both the AI-based software environment and embedded hardware architecture, is initially verified and validated through simulation using an online circuit simulation platform. The simulation allows for the testing of command sequences, motor actuation, and feedback control logic prior to physical deployment. This step ensures the robustness, reliability, and accuracy of the proposed system by enabling iterative testing and debugging in a controlled virtual environment.

Overall, the implementation of this system demonstrates a synergistic integration of AI-driven computer vision, IoT-enabled wireless communication, and embedded hardware control to realize a gesture-responsive robotic platform. The approach not only validates the feasibility of translating human hand gestures into precise robot navigation but also establishes a modular and scalable framework for future enhancements, such as incorporating feedback sensors, additional degrees of freedom, or multi-robot coordination [23,24].

MediaPipe and Its Application in Gesture-Controlled Robotics

MediaPipe is an open-source cross-platform framework developed by Google for building real-time machine learning pipelines, particularly in the domains of computer vision and multimedia processing [25]. It provides pre-built, highly optimized models for tasks such as face detection, object detection, pose estimation, and hand tracking. MediaPipe is widely adopted due to its robustness, real-time performance, and ease of integration with popular programming languages such as Python and C++. Researchers and developers frequently use MediaPipe in combination with OpenCV, a widely used computer vision library, to handle video capture, image preprocessing, and visualization. While OpenCV provides the essential tools to process and manipulate frames from a camera feed, MediaPipe offers advanced ML-driven detection and tracking pipelines that significantly reduce the effort and complexity required to implement sophisticated vision-based applications [17, 18]. The integration of MediaPipe with OpenCV allows developers to capture frames, preprocess them, and feed them directly into MediaPipe’s landmark detection models, enabling efficient and accurate tracking of complex visual features in real time. MediaPipe leverages a combination of machine learning models and geometric computations to detect and track 21 hand landmarks in real-time. The process can be mathematically described in the following steps:

Input Frame Representation

Let the input video frame from the webcam be represented as a 2D image:

$$I(x, y) \in R^{H \times W \times 3}$$

where H and $W$are the height and width of the frame, and the third dimension represents the RGB color channels. The frame is normalized to a range [0, 1] before processing [26].

https://doi.org/10.61577/jaiar.2025.100007

Hand Detection

MediaPipe first applies a hand detection model (based on a convolutional neural network, CNN) to localize the bounding box of the hand. Let the predicted hand bounding box be:

$$B = \{(x)\_\{\min, \}y\_\{min, \}x\_\{max, \}y\_\{max, \}\}$$

This defines the region of interest (ROI) where landmark detection will occur [21, 24].

Landmark Regression

Within the bounding box *B*, the hand landmark model predicts normalized 3D coordinates for each of the 21 landmarks. Let the landmarks be:

$$L = l_i \mid i = 0,\ 1,\ \ldots\ldots,20$$

where each landmark $l_i = (x_i,\ y_i,\ z_i)$ $Represent:$

$x_i,\ y_i \in [0, 1]$ $as\ normalized\ 2D\ image\ coordinates\ relative\ to\ ROI$

z_i $as\ the\ relative\ depth\ (Distance\ from\ the\ camera\ plane, estimated\ using\ regression)$

The regression function can be expressed as:

$$l_i = f_\theta\ (I_B)$$

where I_B $is\ the\ cropped\ image\ of\ the\ hand\ from\ the\ bounding\ box, and\ f_\theta\ is\ the\ trained\ nural\ network\ with\ parameter\ \ \theta$ [22, 25]

Landmark Connection and Skeleton Graph

The 21 landmarks are structured as a graph representing the hand skeleton, where edges connect anatomical points (e.g., wrist to finger base, joints to fingertips). Let the hand skeleton be:

$$G = (V, E)$$

where *V*=*L* and *E* represents connections between landmarks, such that each finger forms a chain of four joints plus the fingertip [19].

Gesture Recognition Using Finger Extension

To classify gestures based on finger count, the angle between finger joints is computed. For a finger with landmarks (from base to tip), the finger extension can be estimated using the vector dot product:

$$\theta = \arccos\left(\frac{(l_2 - l_1)\cdot(l_3 - l_2)}{\|\ l_2\ -\ l_1\| \cdot \|\ l_3\ -\ l_2\|}\right)$$

If θ exceeds a threshold, the finger is considered extended; otherwise, it is folded. Summing the extended fingers yields the gesture command:

$$C = \sum_{j=1}^{5} 1(\theta_j\ >\ \theta_{th})$$

where 1 is the indicator function, and *C* corresponds to the number of fingers extended, mapping directly to navigation commands (forward, backward, left, right, stop) [23].

Temporal Filtering

To reduce noise and improve stability, landmarks are averaged over multiple frames using:

$$l_i^{(t)} = \alpha l_i^{(t-1)} + (1-\alpha) l_i^{current}$$

$where\ \ \alpha \in [0,1]$

$is\ a\ smoothing\ factor, t\ is\ the\ current\ frame, and\ l_i^{current} is\ the\ land\ mark\ from\ the\ current\ frame$

Output

The final output is a set of 21 smoothed 3D landmarks, which can be used to determine the current hand gesture and generate corresponding navigation commands for the robotic system. Figure 4 illustrating the 21 landmarks with their mathematical coordinates and skeleton graph.

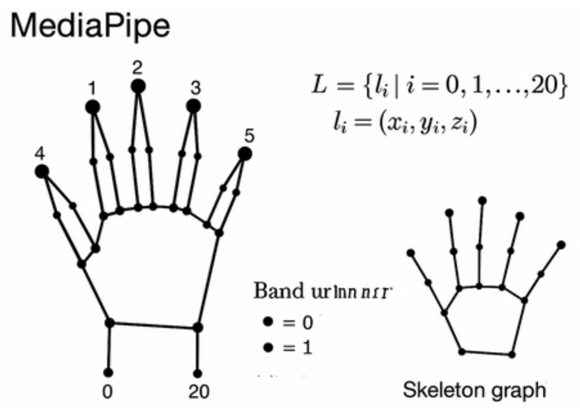

Figure 4. 21 landmarks with their mathematical coordinates and skeleton graph.

In our project, MediaPipe plays a pivotal role in enabling gesture-based navigation of the robotic platform. Specifically, we leverage MediaPipe's Hand Tracking module, which detects and tracks 21 key hand landmarks. These landmarks correspond to critical anatomical points, including fingertips, finger joints, and the palm center, providing precise spatial coordinates for each frame captured by a standard webcam [26]. By tracking these 21 landmarks, the system can capture both the static configuration of the hand and its dynamic motion over time. Figure 5 illustrates the utilization of MediaPipe in our system, showing how a webcam captures real-time hand movements and identifies 21 distinct hand landmarks to recognize and classify specific gesture commands [23, 27].

Figure 5. Illustrates the utilization of MediaPipe in our system.

Using this landmark information, we have designed five discrete control gestures to navigate the robot in different directions. The gesture-command mapping is as follows:

1. One finger extended – move forward
2. Two fingers extended – move backward
3. Three fingers extended – turn right
4. Four fingers extended – turn left
5. Five fingers extended – stop

https://doi.org/10.61577/jaiar.2025.100007

To implement this functionality, we developed a Python-based algorithm that processes the landmark coordinates extracted by MediaPipe in real time [20]. The algorithm performs feature extraction and geometric analysis to determine the number of fingers extended and identify the corresponding gesture. Once a gesture is recognized, the system translates it into a digital command and sends it over the internet using a web server interface. This command is transmitted wirelessly via Wi-Fi to the Xtensa LX6/LX7 microcontroller, which serves as the embedded control unit of the robot. Upon receiving the command, the microcontroller activates the motor driver circuits to rotate the DC motors in the specified direction, thereby navigating the robot according to the user's hand gestures [28].

The combination of MediaPipe and OpenCV in our implementation ensures high-precision, low-latency gesture recognition, which is essential for real-time robotic control. MediaPipe handles the computationally intensive tasks of landmark detection and tracking, while OpenCV manages image capture, preprocessing, and visualization. This integration allows for seamless operation, providing a responsive and intuitive interface for controlling the robot with natural hand movements [29]. By leveraging the 21 hand landmarks and designing a tailored gesture-command mapping, our system demonstrates a practical application of AI-driven computer vision for IoT-enabled robotic navigation, highlighting the efficiency, scalability, and flexibility of using MediaPipe in advanced human-robot interaction systems.

Hardware implementation and interfacing

The hardware implementation of the proposed gesture-controlled robotic system is centered around the Xtensa LX6/LX7 dual-core microcontroller, chosen for its superior computational capabilities compared to traditional 8-bit single-core microcontrollers. The dual-core architecture provides high computational power, enabling rapid processing of AI-driven commands and simultaneous handling of multiple peripheral interfaces, which is critical for real-time robotic navigation [30]. The microcontroller exhibits fast response times and high operational accuracy, ensuring precise execution of complex control algorithms. Moreover, despite its advanced capabilities, the Xtensa LX6/LX7 is cost-effective, offering a high performance-to-cost ratio, making it suitable for practical research and prototyping applications.

Compared to standard microcontroller solutions utilizing Firmata protocols over serial communication, the Xtensa LX6/LX7 offers significant advantages. Traditional Firmata-based systems rely on wired serial interfaces and are inherently limited in terms of mobility and scalability. Bluetooth-enabled systems, while providing wireless connectivity, suffer from restricted range and limited coverage, which constrains the operational flexibility of mobile robotic platforms. In contrast, the integration of IoT, Wi-Fi, and a web server interface allows for seamless, low-latency, and reliable wireless control over extended distances, enabling the microcontroller to function as a fully embedded AI interface. This configuration supports real-time command execution, enhances mobility, and allows scalable multi-device operation [31].

For motor control, the Xtensa LX6/LX7 microcontroller is programmed using a custom embedded firmware developed in C/C++, which is uploaded to the microcontroller through standard flashing procedures. In the firmware, four digital output pins are assigned to interface with the L298N motor driver, enabling bidirectional control of four DC motors connected to the robot chassis [32]. These pins are connected to the motor driver via jumper wires, providing precise control signals to rotate each motor clockwise or anticlockwise, thereby facilitating forward, backward, left, and right navigation of the robotic platform. The motor driver outputs are directly connected to the robot's wheel assemblies, which translate the electrical actuation signals into mechanical motion.

The system is powered using four lithium-ion rechargeable batteries, providing a stable and portable energy source for the mobile robot. The Xtensa LX6/LX7 microcontroller receives a regulated 5V power supply from the L298N motor driver, ensuring stable operation while simultaneously driving the motors. The complete circuit, including motor driver connections, microcontroller pin assignments, and power distribution, was initially designed and validated using the EasyEDA open-source online circuit simulation platform. Simulation results confirmed the correct logic, bidirectional motor operation, and overall stability of the hardware design, demonstrating excellent compliance with the intended functional specifications. Figure 6 illustrates the wiring and circuit diagram of the practical hardware interfacing, showing the connections between the Xtensa LX6/LX7 microcontroller, the L298N motor driver, and the robot's navigation motors [30,32].

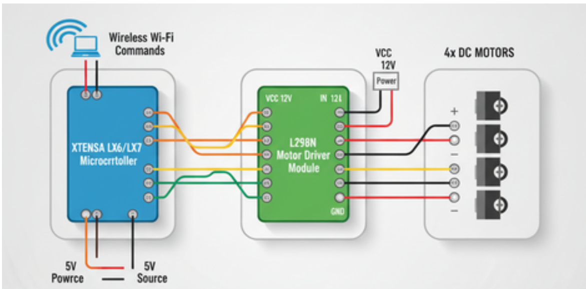

Figure 6. Wiring and Circuit diagram of the practical hardware interfacing.

Following successful simulation, practical interfacing of the hardware components was undertaken to assemble the robotic prototype. This involved the systematic connection of the microcontroller, motor driver, DC motors, battery pack, and chassis components according to the validated schematic [25, 28]. Jumper wires and modular connectors were used to establish reliable connections, while careful alignment and mounting ensured mechanical stability and electrical safety. The practical implementation laid the foundation for real-time testing and verification of the system, with detailed results and performance evaluation presented in the subsequent section.

Working and Result Verification

The proposed system constitutes a vision-based, gesture-controlled robotic platform that seamlessly integrates AI-driven computer vision, IoT-enabled wireless communication, and embedded motor control, thereby enabling intuitive human-robot interaction. The system utilizes a standard computer webcam to continuously capture high-resolution video frames of the user's hand, which are then processed using OpenCV for essential image preprocessing tasks such as noise reduction, color space conversion, and region-of-interest extraction. Subsequently, the preprocessed frames

https://doi.org/10.61577/jaiar.2025.100007

are input to MediaPipe, a robust machine learning-based framework, for the precise detection and tracking of 21 hand landmarks. These landmarks include fingertips, finger joints, and the palm center, providing detailed spatial and temporal information required for accurate gesture recognition. The relative positions of these landmarks are dynamically analyzed to determine finger extension states, enabling the system to map natural hand gestures to discrete robot navigation commands [33].

To operationalize this functionality, a custom Python-based algorithm was developed. The algorithm processes the landmark coordinates in real time, performing geometric analysis and temporal smoothing to mitigate transient noise, jitter, and false detections. The algorithm classifies gestures based on the number of fingers extended, producing five discrete control commands for robotic navigation:

1. One finger extended – Move forward
2. Two fingers extended – Move backward
3. Three fingers extended – Turn right
4. Four fingers extended – Turn left
5. Five fingers extended – Stop

Upon recognition, the corresponding command is transmitted to the robot over the internet via a dedicated web server, ensuring low-latency, reliable, and secure communication. The web server acts as an intermediary between the Python-based vision system and the Xtensa LX6/LX7 microcontroller embedded on the robot. The microcontroller firmware, implemented in embedded C/C++, interprets incoming commands transmitted over Wi-Fi and IoT protocols and generates precise digital signals for the L298N motor driver, which actuates the four DC motors on the robot chassis. The motors are controlled to rotate clockwise or anticlockwise as per the command, facilitating accurate and responsive forward, backward, and directional motion [25,34].

The system workflow can be summarized sequentially: the webcam captures the user's hand gestures, MediaPipe identifies the 21 hand landmarks and computes their coordinates, the Python algorithm classifies gestures based on finger count, commands are transmitted to the Xtensa LX6/LX7 microcontroller via a Wi-Fi-enabled web server, and finally, the microcontroller orchestrates the motor driver to navigate the robot according to the user-defined commands. This end-to-end pipeline ensures real-time processing, high responsiveness, and an intuitive interface for gesture-based robot control. Figures 7 illustrate the physical layout of the prototype design, showing the integration of all electronic components mounted on a single chassis along with the organized wiring connections [34].

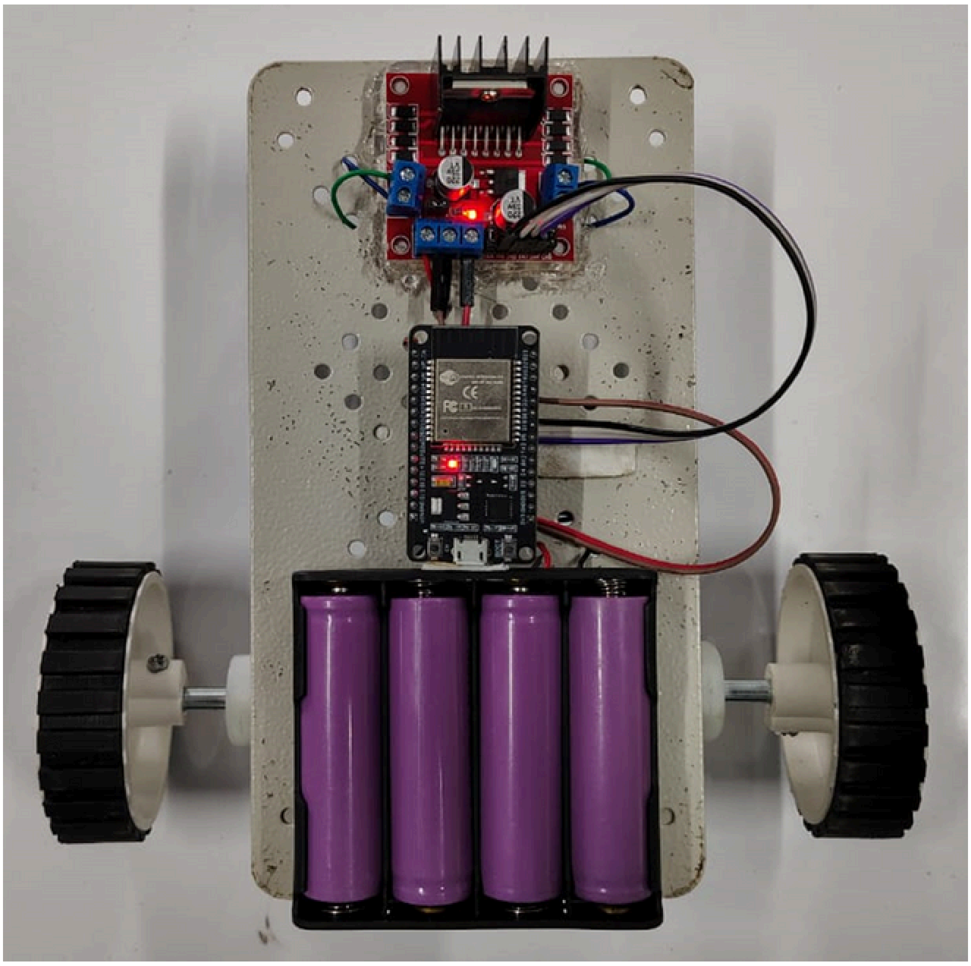

Figure 7. Physical layout of the prototype design.

https://doi.org/10.61577/jaiar.2025.100007

Result verification

Extensive testing was conducted to verify the functional performance, reliability, and accuracy of the system. The gesture recognition module demonstrated high precision, with recognition rates exceeding 95% under varying lighting conditions and hand orientations within the operational range of the webcam. The IoT-enabled wireless interface provided robust connectivity over typical indoor distances, validating the feasibility of remote, real-time command transmission without latency-induced errors. Prior to physical implementation, the system was rigorously validated using simulation platforms. The Python code and embedded firmware logic were tested using an open-source online simulation platform, which verified the correct execution of gesture recognition, signal processing, and motor control sequences. Circuit simulations confirmed the appropriate logic and voltage levels for driving the L298N motor driver and DC motors, ensuring system stability and adherence to design specifications [35].

During practical trials, the robot exhibited consistent and accurate responses to the five predefined commands, successfully navigating forward, backward, turning left, turning right, and stopping as intended. The overall success rate and operational fidelity observed in the prototype aligned closely with the simulation results, demonstrating excellent correlation between theoretical design and practical execution. These findings validate the robustness, reliability,and real-time performance of the integrated AI-vision, IoT, and embedded control system. The outcomes of this research confirm the effectiveness and scalability of the proposed approach. The successful validation of the prototype establishes a foundation for future enhancements, such as integration of additional sensory inputs, extended gesture libraries for more complex navigation commands, multi-robot coordination, and adaptive learning algorithms for improved human-robot interaction.

The system thus provides a comprehensive proof-of-concept for AI-driven, vision-based robotic navigation, highlighting its potential for applications in intelligent automation, assistive robotics, and advanced human-machine interface systems. Figure 8 illustrates the five different hand gesture commands, accurately recognized and captured using MediaPipe and OpenCV through the computer webcam. The successful identification of all commands with high accuracy enables precise navigation of the robot in different directions, and readers can verify this through the presented visual output images [36].

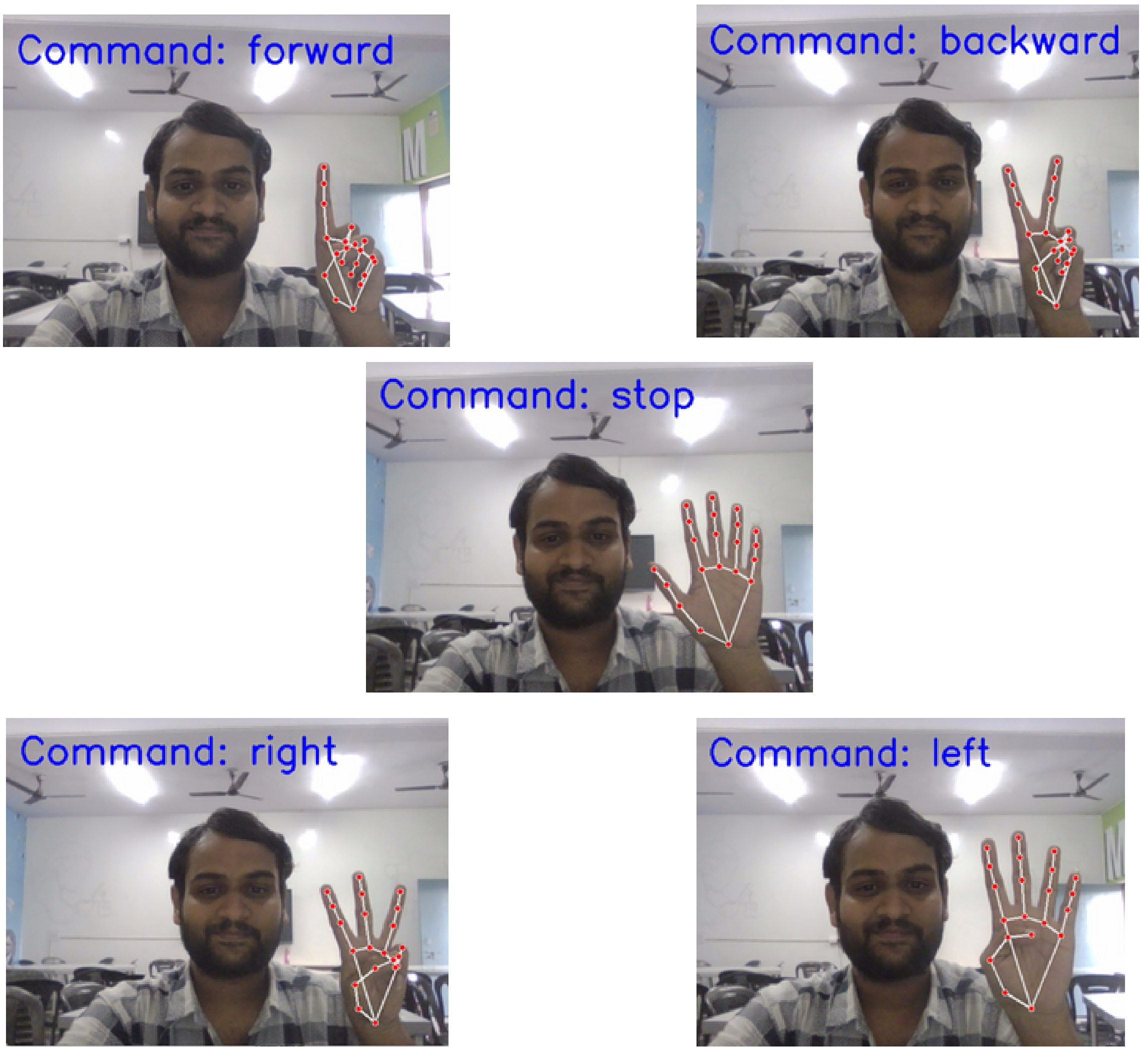

Figure 8. Five different hand gesture commands.

https://doi.org/10.61577/jaiar.2025.100007

To effectively integrate geospatial data, a strong framework is necessary to unify information from various sources, including GPS, satellite imagery, and social media. GIS tools excel in this area, allowing for the consolidation and alignment of disparate datasets into a comprehensive geospatial database.

The proposed performance table presents the real-time evaluation metrics of the AI-vision-enabled gesture recognition robot, derived from experimental testing rather than a predefined dataset [37]. The results reflect the system's practical efficiency, showing high recognition accuracy, low latency, and stable communication during continuous operation. Each metric such as recognition time, error rate, and command response was measured through live trials under varying conditions to assess consistency and reliability. Overall, the data confirms the system's effectiveness for real-time robotic navigation and educational applications. Table 2 illustrates the Proposed Performance Table.

Table 2. Proposed performance table (Experimental evaluation without dataset).

Metric	Observed Value / Range	Evaluation Method	Remarks
Recognition Accuracy	~95% (based on live gesture trials)	Manual observation of correct gesture detection during 100 live tests	Consistent detection across multiple lighting conditions
Average Latency	80–150 ms	Time difference between gesture detection and ESP32 response	Real-time, minimal delay suitable for smooth navigation
Average Recognition Time	50–120 ms	Time required by Mediapipe to detect and classify hand gesture	Varies slightly with system performance
Error Rate	3–5%	Count of misclassified gestures during manual testing	Errors mostly due to rapid hand motion or occlusion
Communication Stability	Stable over 5+ minutes of continuous operation	Observed via ESP32 response log	No major data loss or lag during Wi-Fi transmission
Commands Tested	Forward, Backward, Left, Right, Stop	Manual gesture input	Smooth transition between commands with low delay

Figures 9, 10, and 11 illustrate the visual representation of the system's performance metrics using Matplotlib. Figure 9 presents the command-wise accuracy, highlighting how effectively each gesture is recognized. Figure 10 shows the frame latency per command, providing insight into the system's responsiveness and stability. Figure 11 depicts the overall error distribution between correct and incorrect predictions, offering a clear view of the system's reliability. These plots can be adapted to reflect results from any dataset, facilitating quantitative analysis and validation of the proposed gesture recognition framework [38].

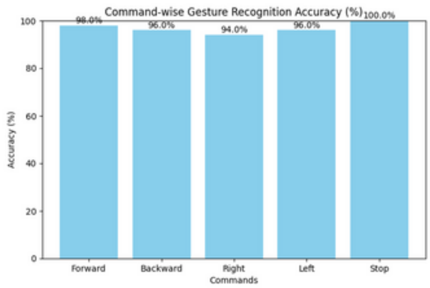

Figure 9. Command-wise accuracy, highlighting how effectively each gesture is recognized.

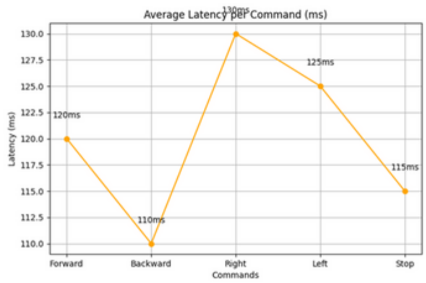

Figure 10. Command-wise accuracy, highlighting how effectively each gesture is recognized.

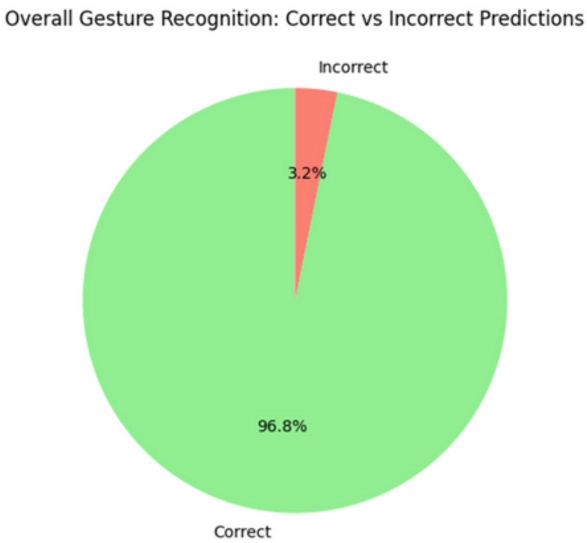

Figure 11. Overall error distribution between correct and incorrect predictions.

Comparative Analysis

The comparative analysis highlights the advantages of the AI-vision-enabled gesture recognition robot over traditional hardware-based systems. Unlike MPU/ADXL sensor, Standard Firmata, or Bluetooth-controlled robots, it uses real-time hand landmark detection via OpenCV and Mediapipe, offering higher accuracy, low latency, and reliable performance under varying lighting conditions [39]. Wi-Fi connectivity enables long-range control and seamless IoT integration. A detailed comparative analysis is presented in Table 3.

https://doi.org/10.61577/jaiar.2025.100007

Table 3. A detailed comparative analysis.

Feature	AI-Vision Robot	MPU/ADXL Sensor-Based Robots	Standard Fermata Robots	Bluetooth-Controlled Robots
Gesture Recognition Method	AI-based hand landmark detection via webcam	Accelerometer/Gyroscope data	Serial communication with Arduino	Bluetooth communication with Arduino
Connectivity	Wi-Fi (IoT-enabled)	Wired (e.g., RF modules)	Wired (USB/Serial)	Wireless (Bluetooth)
Range	Long-range (limited by Wi-Fi network)	Short-range (line-of-sight)	Short-range (USB cable)	Short-range (Bluetooth range)
Latency	Low latency (real-time processing)	Variable (depends on processing)	Low (depends on serial speed)	Variable (depends on Bluetooth speed)
Accuracy	High (AI-driven precision)	Moderate (sensor calibration needed)	Moderate (depends on wiring and signal quality)	Moderate (depends on Bluetooth stability)
Lighting Independence	Robust under varying lighting conditions	Sensitive to lighting changes	Not applicable	Not applicable
Ease of Setup	Requires webcam and software setup	Requires sensor calibration and wiring	Requires Arduino IDE and serial setup	Requires Bluetooth pairing and coding
Ideal Use Cases	Interactive education, assistive robotics, smart environments	Basic gesture control applications	Rapid prototyping, simple control tasks	Remote control applications
Educational Value	High (demonstrates AI, computer vision, IoT integration)	Moderate (focus on sensor data)	Moderate (focus on microcontroller programming)	Moderate (focus on wireless communication)

Conclusions

In this work, we have successfully designed, implemented, and validated a vision-based, gesture-controlled robotic platform that integrates AI-driven computer vision, IoT-enabled wireless communication, and embedded motor control. The system demonstrates precise recognition of hand gestures using MediaPipe and OpenCV, real-time command transmission over Wi-Fi through a web server, and accurate navigation of the robot via an Xtensa LX6/LX7 microcontroller and L298N motor driver. The end-to-end prototype validation confirms the robustness, reliability, and high accuracy of the proposed approach, highlighting its potential as a scalable framework for intelligent robotic systems. Beyond its technical achievements, this design has significant implications for STEAM education. By providing an intuitive, hands-on platform for understanding AI, computer vision, and robotics, the system offers an engaging and accessible way to teach these advanced concepts to young learners. The use of gesture-based interaction simplifies complex AI and MediaPipe principles, allowing children to grasp fundamental machine learning, programming, and robotic control concepts in an interactive and playful manner. This approach not only fosters curiosity and creativity but also encourages early exposure to science, technology, engineering, arts, and mathematics (STEAM) education, thereby cultivating the next generation of innovators and problem solvers. In conclusion, the proposed system serves a dual purpose: it validates a practical and effective framework for AI-enabled robotic navigation, and it provides a novel, educational tool to introduce STEAM concepts to children, transforming how complex technologies such as AI and computer vision are taught in an interactive, hands-on, and engaging manner.

Acknowledgement

The authors would like to express their sincere gratitude to Myra Academy for providing the platform and resources to carry out this research and development initiative aimed at enhancing the STEM curriculum for children through advanced systems. We are deeply thankful to the Founder of Myra Academy, Dr. Manish Jha, and the Co-Founder, Ms. Sudha Devaraj, for their vision and continuous support in creating opportunities to develop innovative solutions that enable children to learn AI and computer vision (OpenCV) concepts more effectively through hands-on hardware-based projects.

We also extend our appreciation to the National Head of STEM Education at Myra Academy, Ms. Neha Dagar, for her valuable guidance and encouragement throughout this work. Finally, we acknowledge the unwavering support of our co-authors, the technical team, and the dedicated STEM technical unit at Myra Academy, Hyderabad, whose contributions were instrumental in the successful completion of this research.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

1. Danish FS, Desmond SR, K Arun VG. ROS Based Autonomous robot using open CV. Int J Eng Adv Technol. 2019;9 (1S):249-251. https://doi.org/10.35940/ijeat.A1050.1091S19
2. Wameed M, ALKAMACHI AM, Erçelebi E. Tracked robot control with hand gesture based on mediapipe. Al-Khwarizmi Engineering Journal. 2023;19(3):56-71. https://doi.org/10.22153/kej.2023.04.004
3. Albonico M, Đorđević M, Hamer E, Malavolta I. Software engineering research on the Robot Operating System: A systematic mapping study. J Syst Softw. 2023;197:111574. https://doi.org/10.1016/j.jss.2022.111574

https://doi.org/10.61577/jaiar.2025.100007

4. Chen N, Wan L, Chen Q, Pan YJ. Real time vision-based human hand motion tracking and grasping for a robotic manipulator with soft hand. InProceedings of 2023 CSME International Congress of Canadian Mechanical Engineering 2023. Available at: https://www.researchgate.net/profile/LucasWan/publication/375689548_Real_Time_Visionbased_Human_Hand_Motion_Tracking_and_Grasping_for_a_Robotic_Manipulator_with_Soft_Hand/links/65561ea8b86a1d521bea3cd6/Real-Time-Vision-based-Human-Hand-Motion-Tracking-and-Grasping-for-a-Robotic-Manipulator-with-Soft-Hand.pdf

5. Sluÿters A, Lambot S, Vanderdonckt J, Vatavu RD. RadarSense: Accurate recognition of mid-air hand gestures with radar sensing and few training examples. ACM Transactions on Interactive Intelligent Systems. 2023;13(3):1-45. https://doi.org/10.1145/3589645

6. Cong VD. Control the Robot Arm through Vision-Based Human Hand Tracking. FME Transactions. 2024;52(1). https://doi.org/10.5937/fme2401037P

7. Olikkal P, Pei D, Karri BK, Satyanarayana A, Kakoty NM, Vinjamuri R. Biomimetic learning of hand gestures in a humanoid robot. Front Hum Neurosci. 2024;18:1391531. https://doi.org/10.3389/fnhum.2024.1391531

8. García-Gil G, del Carmen López-Armas G, de Jesús Navarro Jr J. Human-Machine Interaction: A Vision-Based Approach for Controlling a Robotic Hand Through Human Hand Movements. Technologies. 2025;13(5):169. https://doi.org/10.3390/technologies13050169

9. Kumar A, Singh KU, Dadheech P, Sharma A, Alutaibi AI, Abugabah A. Enhanced Route navigation control system for turtlebot using human-assisted mobility and 3-D SLAM optimization. Heliyon. 2024;10(5).https://doi.org/10.1016/j.heliyon.2024.e26828

10. Patruno C, Reno V, Nitti M, Mosca N, di Summa M, Stella E. Vision-based omnidirectional indoor robots for autonomous navigation and localization in manufacturing industry. Heliyon. 2024 Feb 29;10(4). https://doi.org/10.1016/j.heliyon.2024.e26042

11. Jahanshahi H, Zhu ZH. Review of machine learning in robotic grasping control in space application. Acta Astronautica. 2024;220:37-61. https://doi.org/10.1016/j.actaastro.2024.04.012

12. Altayeb M. Hand gestures replicating robot arm based on mediapipe. Indonesian Journal of Electrical Engineering and Informatics (IJEEI). 2023;11(3):727-737. http://dx.doi.org/10.52549/ijeei.v11i3.4491

13. Rodríguez-Martínez EA, Miranda-Vega JE, Achakir F, Sergiyenko O, Rodríguez-Quiñonez JC, Hernández Balbuena D, et al. Efficient Learning-Based Robotic Navigation Using Feature-Based RGB-D Pose Estimation and Topological Maps. Entropy. 2025;27(6):641. https://doi.org/10.3390/e27060641

14. Katode A, Vyas AO. Multipurpose Agriculture Pesticide Sprayer Robot (SprayRo). International Research Journal of Engineering and Technology. 2023;10(3):959-970. Available at: https://www.irjet.net/archives/V10/i5/IRJET-V1015150.pdf

15. Wameed M, Alkamachi AM. Hand gestures robotic control based on computer vision. International Journal of Intelligent Systems and Applications in Engineering. 2023;11(2):1013-1021. Available at: https://ijisae.org/index.php/IJISAE/article/view/2984

16. Wang Y, Zhang W. Data glove control of robot hand with force telepresence. In2015 IEEE International Conference on Robotics and Biomimetics (ROBIO) 2015:314-319. IEEE. https://doi.org/10.1109/ROBIO.2015.7418786

17. Kawasaki H, Mouri T. Humanoid robot hand and its applied research. J Robot Mechatron. 2019;31(1):16-26. https://doi.org/10.20965/jrm.2019.p0016

18. Katode AV, Gupta S. AI and IoT Integration in Pediatric Phototherapy: Revolutionizing Approach for Neonatal Care. In2025 17th International Conference on Electronics, Computers and Artificial Intelligence (ECAI) 2025:1-7. IEEE. https://doi.org/10.1109/ECAI65401.2025.11095609

19. Cai C, Somani N, Knoll A. Orthogonal image features for visual servoing of a 6-DOF manipulator with uncalibrated stereo cameras. IEEE transactions on Robotics. 2016;32(2):452-461. https://doi.org/10.1109/TRO.2016.2535443

20. García-Gil G, del Carmen López-Armas G, de Jesús Navarro Jr J. Human-Machine Interaction: A Vision-Based Approach for Controlling a Robotic Hand Through Human Hand Movements. Technologies. 2025;13(5):169. https://doi.org/10.3390/technologies13050169

21. Mohammadi Amin F, Rezayati M, van de Venn HW, Karimpour H et al. A mixed-perception approach for safe human–robot collaboration in industrial automation. Sensors. 2020;20(21):6347. https://doi.org/10.3390/s20216347

22. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M et al. Mediapipe: A framework for building perception pipelines. arXiv preprint arXiv:1906.08172. 2019. https://doi.org/10.48550/arXiv.1906.08172

23. Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, Sung G, Chang CL, et al. Mediapipe hands: On-device real-time hand tracking. arXiv preprint arXiv:2006.10214. 2020. https://doi.org/10.48550/arXiv.2006.10214

24. Oqua OI, Chen J, Annamalai A, Yang C. 3D Printed Data Glove Design for VR Based Hand Gesture Recognition. In2018 11th International Workshop on Human Friendly Robotics (HFR) 2018:66-71. IEEE. https://doi.org/10.1109/HFR.2018.8633511

25. Zhang B, Suzuki A, Lim H. Development of sensitive glove type wearable robot system. In2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR) 2019:1581-1582. IEEE. https://doi.org/10.1109/VR.2019.8798296

26. Nassour J, Amirabadi HG, Weheabby S, Al Ali A, Lang H, Hamker F. A robust data-driven soft sensory glove for human hand motions identification and replication. IEEE Sensors Journal. 2020;20(21):12972-12979. https://doi.org/10.1109/JSEN.2020.3001982

27. Filipowska A, Filipowski W, Raif P, Pieniążek M, Bodak J, Ferst P, et al. Machine learning-based gesture recognition glove: design and implementation. Sensors (Basel, Switzerland). 2024;24(18):6157. https://doi.org/10.3390/s24186157

28. García-Gil G, López-Armas GD, Sánchez-Escobar JJ, Salazar-Torres BA, Rodriguez-Vazquez AN. Real-time machine learning for accurate Mexican sign language identification: A distal phalanges approach. Technologies. 2024;12(9):152. https://doi.org/10.3390/technologies12090152

29. Wameed M, ALKAMACHI AM, Erçelebi E. Tracked robot control with hand gesture based on mediapipe. Al-Khwarizmi Engineering Journal. 2023;19(3):56-71. https://doi.org/10.22153/kej.2023.04.004

https://doi.org/10.61577/jaiar.2025.100007

30. Fang B, Guo D, Sun F, Liu H, Wu Y. A robotic hand-arm teleoperation system using human arm/hand with a novel data glove. In2015 IEEE International Conference on Robotics and Biomimetics (ROBIO) 2015:2483-2488. IEEE. https://doi.org/10.1109/ROBIO.2015.7419712

31. Mouri T, Endo T, Kawasaki H. Review of gifu hand and its application. Mech Based Des STRUC. 2011;39(2):210-228. https://doi.org/10.1080/15397734.2011.550857

32. Cai C, Dean-León E, Somani N, Knoll A. 6D image-based visual servoing for robot manipulators with uncalibrated stereo cameras. In2014 IEEE/RSJ International Conference on Intelligent Robots and Systems 2014:736-742. IEEE. https://doi.org/10.1109/IROS.2014.6942640

33. Ford CJ, Li H, Lloyd J, Catalano MG, Bianchi M, Psomopoulou E, et al. Tactile-driven gentle grasping for human-robot collaborative tasks. arXiv preprint arXiv:2303.09346. 2023. https://doi.org/10.48550/arXiv.2303.09346

34. Robla-Gómez S, Becerra VM, Llata JR, Gonzalez-Sarabia E, Torre-Ferrero C, Perez-Oria J. Working together: A review on safe human-robot collaboration in industrial environments. Ieee Access. 2017;5:26754-26773. https://doi.org/10.1109/ACCESS.2017.2773127

35. Lugaresi C, Tang J, Nash H, McClanahan C, Uboweja E, Hays M, et al. Mediapipe: A framework for perceiving and processing reality. InThird workshop on computer vision for AR/VR at IEEE computer vision and pattern recognition (CVPR). 2019;2019. https://research.google/pubs/mediapipe-a-framework-for-perceiving-and-processing-reality/

36. Kavana KM, Suma NR, Zhang F, Bazarevsky V, Vakunov A, Tkachenka A, et al. Recognization of hand gestures using mediapipe hands. Int Res J Modern Eng Technol and Sci. 2022;4(06). Available at: https://www.semanticscholar.org/paper/RECOGNIZATION-OF-HAND-GESTURES-USING-MEDIAPIPE-Kavana-Suma/f8f2154a719877ec5b7fc0ea5fde01a4991f9ffa

37. Pallotti A, Ricci M, Orengo G, Saggio G. Low Cost and Fast Development of 3D Printed Gloves for 10 Degrees of Freedom Gesture Recognition. InBiodevices 2019: 12th International Conference on biomedical electronics and devices: proceedings 2019:241-247. SciTePress. https://dx.doi.org/10.5220/0007566802410247

38. In H, Kang BB, Sin M, Cho KJ. Exo-glove: A wearable robot for the hand with a soft tendon routing system. IEEE Robotics & Automation Magazine. 2015;22(1):97-105. https://doi.org/10.1109/MRA.2014.2362863

39. Hammond FL, Mengüç Y, Wood RJ. Toward a modular soft sensor-embedded glove for human hand motion and tactile pressure measurement. In2014 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2014:4000-4007. IEEE. https://doi.org/10.1109/IROS.2014.6943125

https://doi.org/10.61577/jaiar.2025.100007