MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 500-505

Indoor Surveillance in Autonomous Ground Vehicle using AI & IoT

[1]Baby Priyadharshini P*, [2]Ananth S, [3]Gomathi M, [4]Elamvendhan K,
[5]Rishikumar K, [6]Sri Nath R, [7]Surya N
[1,3]Assistant Professors, [2]Associate Professor, [4,5,6,7]UG Scholars
babypriyadharshinip@mahendra.info[1], anand.siet@gmail.com[2], gomathim@mahendra.info[3], elamvendhank@gmail.com[4], rishikannan@gmail.com[5], srinath25k@gmail.com[6], suryanagarajan515@gmail.com[7]
Department of Artificial Intelligence and Data Science, Mahendra Engineering College(Autonomous), Namakkal, Tamil Nadu, India.

ABSTRACT

Modern indoor security systems often rely on static camera networks, which suffer from blind spots, or expensive, or else fully autonomous robotic platforms that require heavy onboard processing capabilities. The development of a dual-mode (indoor) Autonomous Ground Vehicle, an indoor surveillance device, has been demonstrated in this paper using distributed edge-cloud architecture. The physical edge device consists of an ESP32 microcontroller which performs the following functions using 4 sensors: localization - HC-SR04 Ultrasonic sensors for obstacle avoidance, environmental monitoring - MQ-series gas sensors, NTC thermistor for temperature sensors and custom IR receiver for fire detection. The video data produced by the edge device is streamed to a local centralized server to compensate for the edge device's processing limitations and the server runs YOLOv5x for real-time object and threat detection. The system sends an alert to the user via WhatsApp containing a timestamped image once a threat has been detected and triggers an API on a web-based dashboard. This technological demonstration bridges the technological divide between inexpensive Internet-of-Things hardware and high-end computer vision technology by enabling cost-effective and scalable surveillance that can be built and enhanced with new features in a short period of time.

Keywords: Autonomous Ground Vehicle, Artificial Intelligence, Internet of Things, Indoor Surveillance, Deep Learning, Object Detection

1.0 INTRODUCTION

In the last few years, the realms of physical security and surveillance have experienced tremendous change as a result of the rapid merging of Artificial Intelligence and the Internet of Things. Historically, stationary surveillance utilized an arrangement of fixed camera networks primarily designed for indoor applications; while such a system is generally effective for general-purpose monitoring alone, static cameras are subject to limitations such as being unable to provide adequate coverage for all areas due to their inability to capture an entire area without overlapping camera views hence requiring a lot of different cameras to capture any one area [1]. Due to these spatial limitations, mobile surveillance solutions have begun to gain traction with the introduction of Autonomous Ground Vehicles as the new alternative; however, the addition of AGVs to create continuous mobile surveillance poses additional engineering problems, as a new industry is formed from the balance between providing high fidelity computer vision in a mobile environment while at the same time addressing the significant power and computing needs of the mobile hardware [2].

Deep learning advances, particularly single-stage object detection algorithms such as the You Only Look Once framework, have changed the way we identify real-time threats [3]. The YOLO algorithms can quickly and accurately identify particular objects like unauthorized personnel or weaponry [4]. However, the cost and power of the microcomputer needed to run complex algorithms such as YOLOv5x natively on an autonomous ground vehicle make fully autonomous, heavily equipped robots impractical for scalable indoor installations or budget-conscious prototyping [5].

To resolve the processing limitations caused by running high-performance artificial intelligence workloads on the AGV alone, this paper discusses the use of a highly efficient, distributed edge-cloud architectural system. Instead of forcing the mobile robot to run heavy AI workloads locally, the AGV can divide responsibility for operational functions. The AGV is the physical edge device and will use a lightweight ESP32 microcontroller to perform immediate environmental sensing and navigating. It employs a distributed array of sensors including an ultrasonic sensor, NTC thermistor, gas sensor, and custom IR fire detector to monitor its surround and to detect obstacles [6].

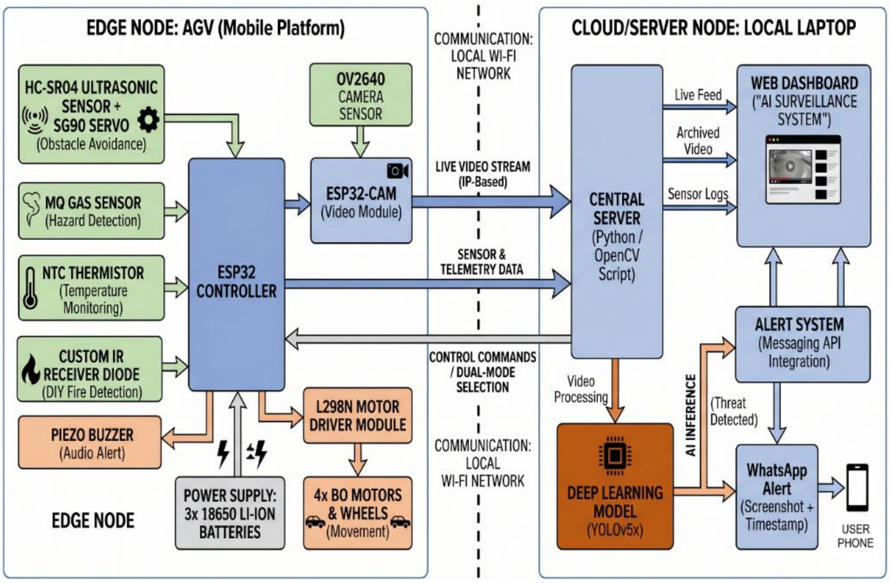

Figure 1.1: Conceptual Architecture Diagram

The onboard ESP32-CAM will simply be a data gathering device that provides live-streaming video frames over a local WI-FI network to a consolidating server. All of the computational burden will be occurring on the consolidating server performing YOLO v5x Object Detection in real-time for any possible threat, as well integrating Internet-of-Things protocols through a web dashboard that will flag any threats and will also provide time-stamped screenshots of any threats whenever an automated WhatsApp alert is sent to the user; therefore this system will provide a very light weight, rapid response and dependable security system through the separation of physical navigation and AI inference processing.

2.0 HARDWARE ARCHITECTURE AND EDGE IMPLEMENTATION

The purpose of this project is to develop an Autonomous Ground Vehicle capable of conducting indoor surveillance missions. The vehicle's physical structure needs to have a balance between structural mobility, power delivery over an extended period of time, and the integration of the vehicle's sensors into one seamless unit [8]. The AGV's hardware structure is built to be modular, which allows for the separation of core functions such as locomotion, environmental sensing, and heavy visual processing. The modularity of the AGV allows for an economical solution to the problem of providing a robust performing vehicle.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 500-505

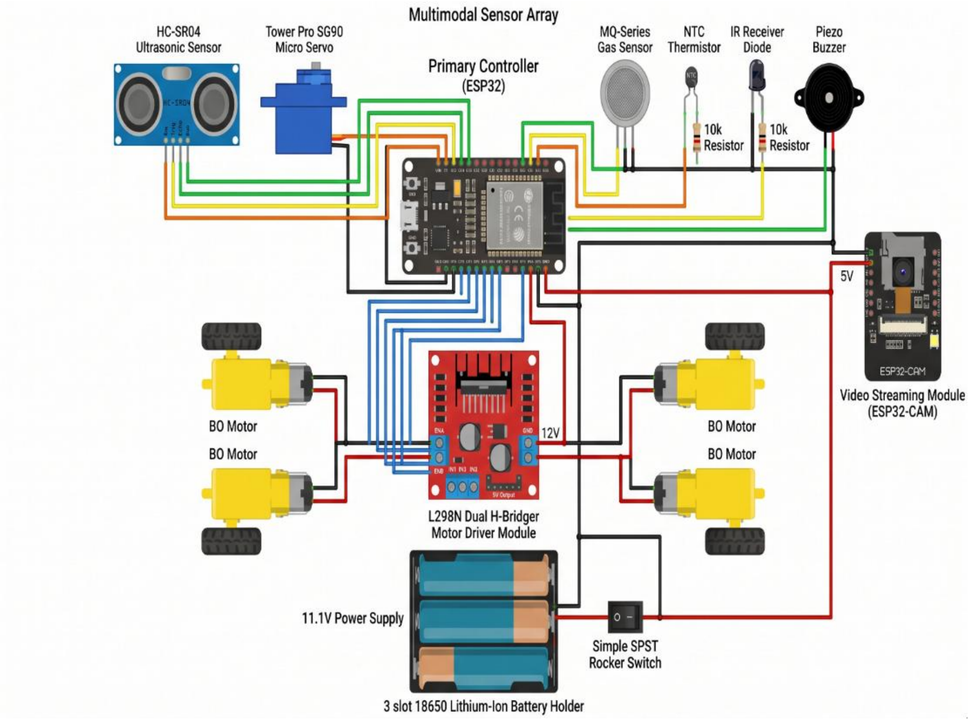

Figure 2.1: The AGV Hardware Circuit Diagram

2.1 STRUCTURAL DESIGN AND LOCOMOTION

AGV's creation is mainly based on an exclusively made two-layer acrylic base, which is kept apart by brass spacers. This vertical structure allows to isolate the lower deck with powerful electric motor circuitry from the upper deck with microcontrollers and environmental sensors therefore reducing electromagnetic interference, [9].Motion of AGVs is accomplished using four driven wheels via 4WD configuration with a set of four DC motor gearboxes - BO motors with rubber wheels and are operated by using a dedicated L298N Dual H-Bridges as a driving module. L298N provides the ability to independently control the speed and direction of the left and right wheel sets, which allows for differential steering capabilities. Low speed high torque drive train is ideal for navigating tight hallways and dynamic obstacles found inside buildings where surveillance is required. [10] Additionally, with the ability to patrolling for extended periods of time without needing to be connected to a power source, one power bank will power the entire AGV edge platform using three 3000mAh 18650 lithium-ion batteries rated at 3.7 volts, connected to a centre-located "single-pole, single-throw" rocker switch providing a stable power source to the entire unit. [11].

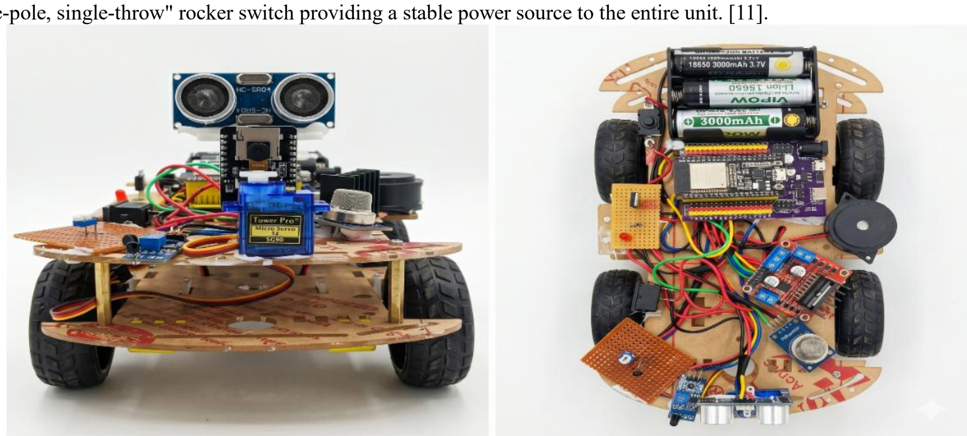

Figure 2.2: Photograph of the Physical Prototype

2.2 PRIMARY EDGE CONTROLLER AND SENSOR ARRAY

he primary edge intelligence is controlled by an ESP32 Development Board. The board was selected for its ability to support high-performance dual-core processing systems as well as its native IoT capabilities. The board serves as the central nervous system to execute local processing and has the function of collecting sensor data from a multi-mode distributed sensor array that continuously samples the physical environment [12]. The AGV uses an HC-SR04 ultrasonic transducer mounted on the front of the chassis using a Tower Pro SG90 micro servo motor for navigation and spatial awareness. The servo sweeps the ultrasonic transducer through a 180-degree field of view and provides the ESP32 with real-time distance measurements and detection of physical objects to facilitate and inhibit tire trajectory correction and obstacle avoidance [13].The AGV also serves as a full-featured environmental monitoring station beyond basic navigation. The sensor array includes an MQ-series gas sensor for detection of airborne hazardous particulates or smoke; an NTC thermistor for ambient temperature monitoring; and a custom-designed infrared receiver diode circuit functioning as a highly sensitive fire detection sensor. The ESP32 can use the locally processed environmental metrics to activate an onboard piezo buzzer to provide audible warning before sending hazard data to the central network [14].

2.3 DEDICATED VIDEO STREAMING MODULE

The architecture of this system purposely separates the camera unit from the main navigation control unit, which is an important design decision [15]. Therefore, the camera is mounted on a separate ESP32-CAM microcontroller with an OV2640 camera sensor for image acquisition.

Due to significant memory and processing limitations associated with edge-level microcontrollers, raw images cannot be processed locally, as this would result in immediate memory overflow/overwhelm and processing limitations of the edge-level microcontrollers. Instead, the ESP32-CAM only delivers video footage captured from the camera as a stream in real-time to the local Wi-Fi network [16]. One of the primary benefits of streaming instead of processing the images locally is the reduced thermal and power loads placed on the vehicle. In addition, since the raw video feed from the camera is being continuously sent to a centralized server for processing, the processing load from the deep learning algorithms that use the camera feed is offloaded from the camera, enabling the camera to operate at its highest efficiency under varying indoor brightness conditions and without delays [17].

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 500-505

3.0 SOFTWARE ARCHITECTURE AND IOT INTEGRATION

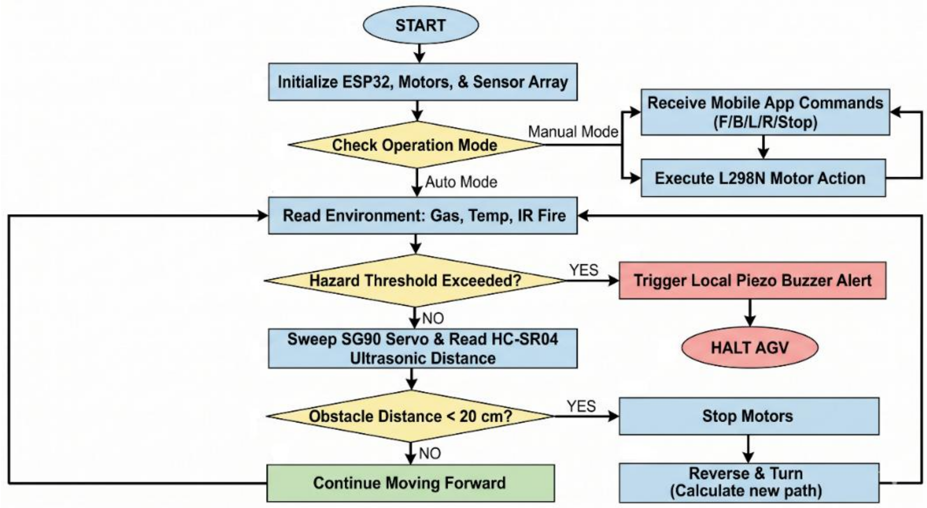

Figure 3.1: Control Logic Flowchart

The edge device that is both highly capable and has a robust software ecosystem will connect the world of physical navigation with the realm of human oversight to control the AGV's navigation. The system's architecture utilized a dual-interface approach to support this purpose utilizing a localized mobile application for immediate teleoperation where the functionality accessible from a mobile device and a centralized web dashboard for comprehensive surveillance monitoring. This dual-interface approach also results in high reliability throughout the wireless network. [18].

3.1 DUAL-MODE NAVIGATION AND MOBILE INTERFACE

The AGV has two modes of navigation: manual and autonomous, allowing it to work in various operational situations. In autonomous mode, the ESP32 is capable of creating a path for itself and avoiding obstacles using data collected from ultrasonic sensors [19]. However, since many indoor environments are complicated and require human assistance to navigate, the AGV has been equipped with an Android-based custom mobile application that allows the operator to take control when necessary.This mobile application connects via local Wi-Fi/Bluetooth to the Edge Device and has a basic control layout (Forward, Backward, Left, Right, Stop) and two toggle buttons to switch from "Auto Mode" to "Manual Mode". High-speed telemetry can experience problems such as latency due to motion feedback; however, the slow movement of the indoor surveillance AGV will not be prone to these types of problems, therefore the ability to directly teleoperate will be very stable and responsive [20]. Finally, the application will provide a real-time view of local environmental data (temperature, gas levels, system status) sent by the AGV's various sensors to the user's display screen.

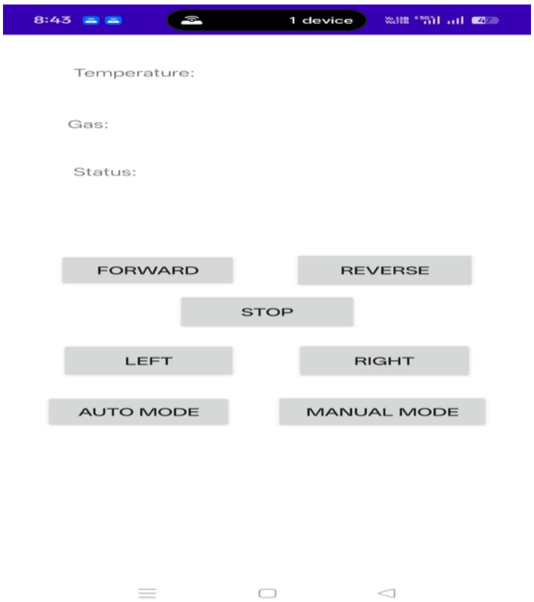

Figure 3.2: Mobile App Interface

3.2 CENTRALIZED WEB DASHBOARD

The data from the mobile app that is associated with the localized physical control is relayed to the central location, AI SURVEILLANCE SYSTEM through a custom web-based portal. The reported data is accessible from the portal via the specific IP of the ESP32-CAM that routes the live video stream. Furthermore, the portal also archives past session data in an organized manner. The past sessions can be categorized by utilizing sub-menus for "Recorded Videos", "Recorded Audio", and "Screenshot". A central logging system for patrol data can ensure that all patrol data is stored off-board and there is no risk of losing the data in the event that the AGV has hardware damage.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 500-505

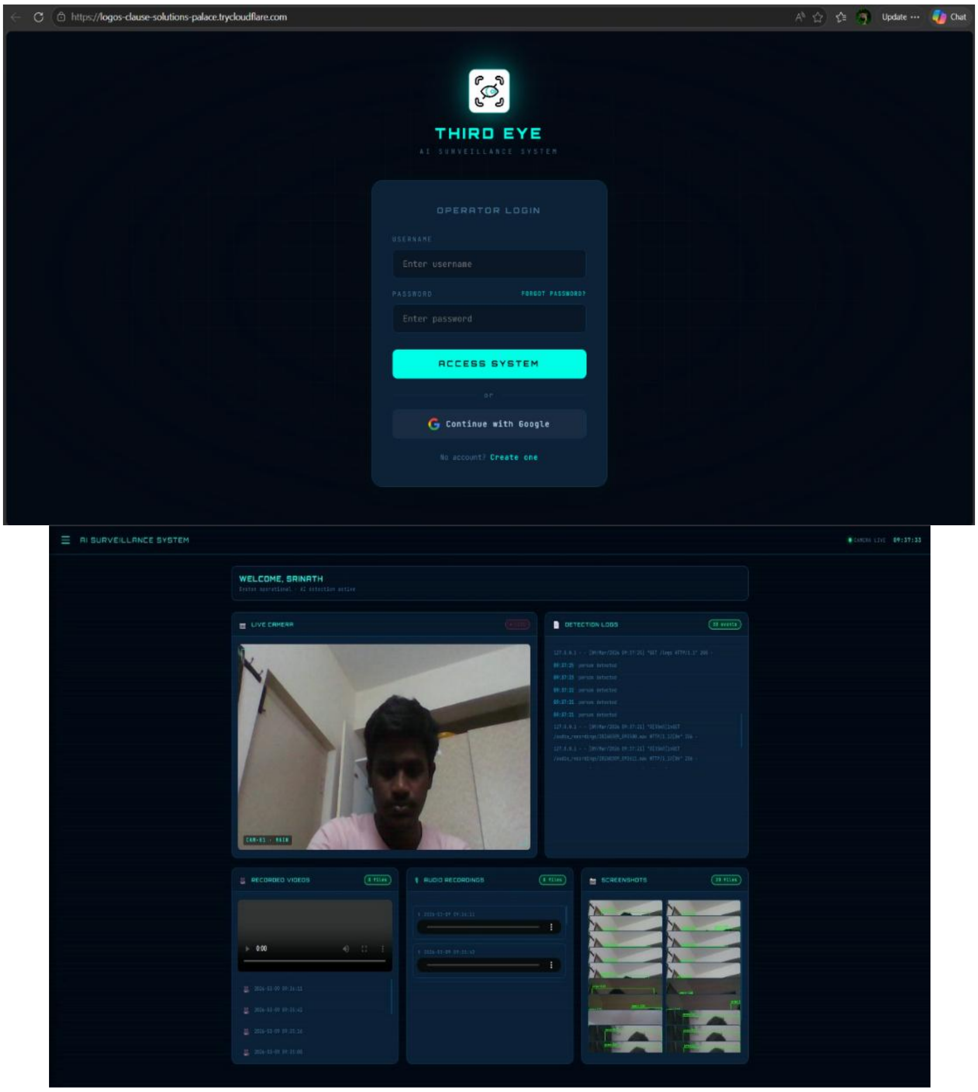

Figure 3.3: "AI SURVEILLANCE SYSTEM" Web Dashboard

4.0 DEEP LEARNING AND THREAT DETECTION

The architecture's most intensive processing phase occurs during the real-time identification of security threats. Since the device is streaming only using the ESP32-CAM as a camera only, all intensive processing requiring significant CPU cycles is completed via an external local computer or a laptop using a Python-based deep learning program script.

4.1 YOLOv5x INTEGRATION AND OBJECT RECOGNITION

The central server uses the OpenCV library to process raw video frames captured from the ESP32-CAM's IP stream. The You Only Look Once (YOLO) framework [22] is employed to help isolate the dynamic threats within a static indoor environment and identify them. YOLOv5x was chosen for its superior accuracy among all the YOLO models available. YOLO is classified as a one-stage object detection system, and it analyzes the entirety of an image frame in one process, where it will output bounding boxes with their class probabilities. This type of object detection system is the most desirable for dynamic video surveillance, as it has a high degree of precision and speed required for real-time monitoring [23].

Figure 4.1: YOLOv5x Accuracy & Performance Report

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 2, 2026, Pages 500-505

The model is programmed to identify various unauthorized actors or hazardous elements in the indoor space (e.g., individuals, knives), and as the target is within the camera view, the YOLO model will make a prediction of the target based on its trained data. If the model determines that the detection confidence exceeds the preset safety threshold, the OpenCV script will overlay a highlighted bounding box on the processed video and enter the corresponding label [24].

Figure 4.2: Real-Time YOLOv5x Detection

4.2 AUTOMATED API ALERT SYSTEM

The ability to detect a threat is only helpful if the information can be communicated quickly and efficiently. Automated messaging protocols were implemented for this reason. When a critical threat is detected, the system will not only log the event on the Web Dashboard, but it will also take a screenshot of the bounding boxes at the time of detection. Once captured, the system uses an integrated Messaging Application Programming Interface to automatically send an emergency message immediately to the user's WhatsApp account. This message will have the timestamped photograph of the threat's classification (e.g., "THREAT: KNIFE") and an indication that the security system has been compromised. This automated notification loop provides for immediate and effective cooperation between security personnel and allows for the rapid response by all necessary parties involved in the emergency [25].

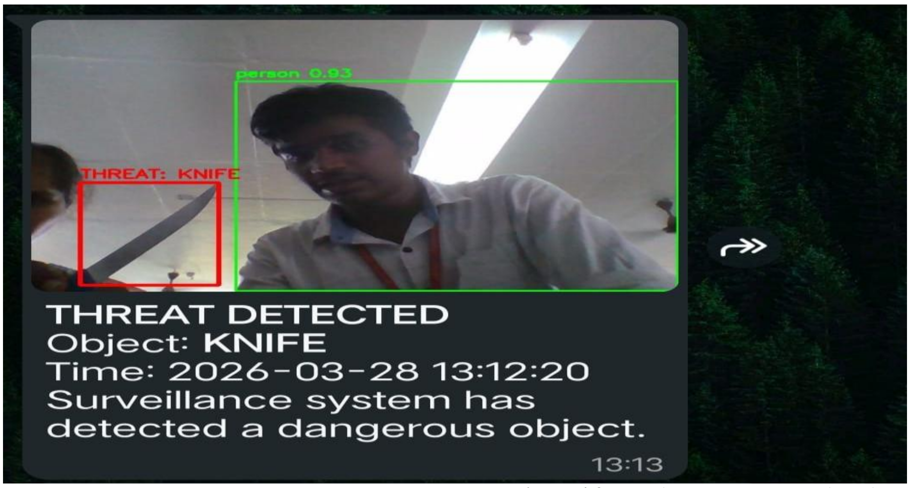

Figure 4.3: The Automated WhatsApp Alert

5.0 RESULT

The Edge-Cloud Physical Deployment of an Autonomous Ground Vehicle used for Surveillance Inside of Building Spaces successfully achieved its objectives of providing indoor surveillance in multiple tests. The ESP32 Device located at the edge offloaded the majority of its computation to a centralized local server, thereby exhibiting superior battery life (power consumption) and navigational stability while generously moving through narrow corridors as well as gathering environmental information like fire, temperature, along with gas measurements using an array of different sensors, without successfully exhausting the memory of the microcontroller.

The ESP32-CAM was used to test visual threat detection capabilities, where it streamed real-time video frames over the local Wi-Fi network with low latency. The central local server used deep learning model YOLOv5x, to process the streamed video frames to identify several different targeted objects, including people carrying harmful and restricted weapons, using previously defined object classifications. Automated API integration exceeded the confidence level programmed into the system, capturing the image frame corresponding to the detected object, overlaid a bounding box on the object and the person, and sent the timestamped still image of the detected person with that object to designated WhatsApp accounts in real-time.

6.0 CONCLUSION

This article discusses the development of a low-cost, dual-mode, autonomous ground vehicle that addresses the need for more affordable Internet of Things devices to be used alongside expensive 'cutting-edge' technology such as computer vision systems. This new type of AGV will use a distributed edge-cloud architecture; thus, decoupling the navigation of the AGV from the AI-related decisions being made, thereby eliminating the need for costly, heat-generating micro-computers on the AGV. The combination of an ESP32-based environmental sensor array, a teleoperation dashboard, and a computer vision-based real-time threat detection system using the YOLO algorithm demonstrates that an efficient and highly automated indoor surveillance system is achievable. As a demonstration of the technology's capability, the prototype created here will provide a basis for the development of higher-capacity AGVs for future commercial and industrial security applications, where speed, scalability and a limited budget are key considerations.

REFERENCES

[1] Wu H, Zeng Q, Guo C, Zhao T, Chen CW. Target-Aware Camera Placement for Large-Scale Video Surveillance. IEEE Transactions on Circuits and Systems for Video Technology. 2024; 34(12):13338-13350.
[2] Noh D, Sung C, Uhm T, et al. X-MAS: Extremely Large-Scale Multi-Modal Sensor Dataset for Outdoor Surveillance in Real Environments. IEEE Robotics and Automation Letters. 2023; 8(2):1093-1100.
[3] Gholami R, Jahromi HD, Sedaghat S. Design and Implementation of a Near Real-Time Human Detection Robot Using YOLO Framework and IoT Technologies. IEEE Access. 2025; 13:3618302.
[4] Wei J, As'arry A, Rezali KAM, Yusoff MZM, Ma H, Zhang K. A Review of YOLO Algorithm and Its Applications in Autonomous Driving Object Detection. IEEE Access. 2025; 13:93711.
[5] Mittal R, Khandelwal D, Nemade KH, Rani G, Dhaka VS, Pathak V. 3D-Printed AI-Powered Robot with Secure Data Handling and Return to Home Capabilities. IEEE Access. 2025; 13:189925.
[6] Zhang L, Hu M, Zhang H, Bian Y, Nguyen A-T, Ding R. Channel-Level Event-Triggered Communication Scheme for Path Tracking Control of Autonomous Ground Vehicles with Distributed Sensors. IEEE Transactions on Vehicular Technology. 2023; 72(10):12553-12565.
[7] Poonia RC, Upreti K, Singh N, Kesarwani J, Alam MS. An Efficient Detection of Suspicious Objects from Dynamic Video Surveillance by Fusion-based Multiview Deep Learning Techniques. Journal of Cyber Security and Mobility. 2024; 13(4):25-45.
[8] Roche J, De-Silva V, Kondoz A. A Multimodal Perception-Driven Self Evolving Autonomous Ground Vehicle. IEEE Transactions on Cybernetics. 2022; 52(9):9279-9290.
[9] Cheng L, Cao S. Radar-Camera Fused Multi-Object Tracking: Online Calibration and Common Feature. IEEE Transactions on Intelligent Transportation Systems. 2026; 27(1):1295-1308.
[10] Jayalakshmi KP, Nair VG, Sathish D. A Comprehensive Survey on Coverage Path Planning for Mobile Robots in Dynamic Environments. IEEE Access. 2025; 13:60185.
[11] Aggravi M, Sirignano G, Giordano PR, Pacchierotti C. Decentralized Control of a Heterogeneous Human-Robot Team for Exploration and Patrolling. IEEE Transactions on Automation Science and Engineering. 2022; 19(4):3109-3122.
[12] Li G, Fan W, Xie H, Qu X. Detection of Road Objects Based on Camera Sensors for Autonomous Driving in Various Traffic Situations. IEEE Sensors Journal. 2022; 22(24):24253-24262.
[13] Tong K, Steinberger M, Horn M, Solmaz S, Watzenig D. Anytime Optimal Trajectory Repairing for Autonomous Vehicles. IEEE Open Journal of Intelligent Transportation Systems. 2025; 6:553.
[14] Wang T, Huang P, Dong G, Zhao Y. A Diffusion-Based Reactive Approach to Road Network Cooperative Persistent Surveillance. IEEE Transactions on Intelligent Transportation Systems. 2024; 25(5):4265-4275.
[15] Suresh MSS, Menon V. A Generic and Scalable Approach to Maximize Coverage in Diverse Indoor and Outdoor Multicamera Surveillance Scenarios. IEEE Transactions on Systems, Man, and Cybernetics: Systems. 2023; 53(2):1172-1185.
[16] Yu Y, Lee S. Remote Driving Control with Real-Time Video Streaming Over Wireless Networks: Design and Evaluation. IEEE Access. 2022; 10:3183758.
[17] Poluan SER, Chen Y-A. Generating Social-Aware Locations for a Robot in a Human Group by Image Generation Using Adversarial Learning. IEEE Access. 2024; 12:3415708.
[18] Li Q, Chang B, Li M, Mei W, Chen Z. A Novel Communication and Control Co-Design Method for Wireless Control Systems: A Communication Perspective. IEEE Internet of Things Journal. 2025; 12(16):33858-33870.
[19] Zhao L, Nybacka M, Aramrattana M, Rothhämel M, Habibovic A, Drugge L. Remote Driving of Road Vehicles: A Survey of Driving Feedback, Latency, Support Control, and Real Applications. IEEE Transactions on Intelligent Vehicles. 2024; 9(10):6086-6102.
[20] Schippers E, Schrank A, Kotian V, Messiou C, Oehl M, Papaioannou G. A Motion for No Motion: The Redundancy of Motion Feedback in Low-Velocity Remote Driving of a Real Vehicle. IEEE Access. 2025; 13:181914.
[21] Lee S. An Efficient Coverage Area Re-Assignment Strategy for Multi-Robot Long-Term Surveillance. IEEE Access. 2023; 11:3260622.
[22] Kim H-J, Shin M-C, Han M-W, Hong C-p, Lee H-W. An Efficient Scheme to Obtain Background Image in Video for YOLO-based Static Object Recognition. Journal of Web Engineering. 2022; 21(6):1705-1724.
[23] Priadana A, Nguyen D-LN, Vo X-T, Choi J, Ashraf R, Jo K. HFD-YOLO: Improved YOLO Network Using Efficient Attention Modules for Real-Time One-Stage Human Fall Detection. IEEE Access. 2025; 13:3547360.
[24] Zhang X, Shu C, Li S, Wu C, Liu Z. AGVS: A New Change Detection Dataset for Airport Ground Video Surveillance. IEEE Transactions on Intelligent Transportation Systems. 2022; 23(11):20588-20600.
[25] Zhang J, Wu Y, Zhou M. Cooperative Dual-Task Path Planning for Persistent Surveillance and Emergency Handling by Multiple Unmanned Ground Vehicles. IEEE Transactions on Intelligent Transportation Systems. 2024; 25(11):16288-16301.