

# IoT-Based Face Recognition AI Robot for Intelligent Surveillance

Rakshitha N B<sup>1</sup>, Rakshitha S<sup>2</sup>, Sanjitha S<sup>3</sup>, Yashaswini H M<sup>4</sup>, Smitha M M<sup>5</sup>

Student, Electronics and Communication Engineering<sup>1-4</sup>

Assistant Professor, Electronics and Communication Engineering<sup>5</sup>

Kalpataru Institute of Technology, Tiptur, India

rakshithanb36@gmail.com, rakshithas614@gmail.com, ssanjitha3@gmail.com,

yashumallikarjun8@gmail.com

**Abstract:** *With the rapid advancement of Artificial Intelligence (AI), the Internet of Things (IoT), and embedded systems, intelligent surveillance solutions have become increasingly important. This paper presents the design and implementation of an IoT-based face recognition AI robot for intelligent surveillance and human-robot interaction. The system integrates ESP32-CAM for real-time image acquisition, Python-based AI algorithms for face detection and recognition, and Arduino-controlled robotics for motion and interaction. Facial landmark-based encoding is employed to achieve reliable recognition with low computational overhead. Experimental results demonstrate recognition accuracy exceeding 90% with response times below two seconds, validating the suitability of the proposed system for real-time embedded surveillance applications.*

**Keywords:** Face Recognition, Internet of Things, ESP32-CAM, Robotics, Embedded Systems, Surveillance

## I. INTRODUCTION

Rapid urbanization has increased the demand for intelligent and autonomous surveillance systems. Traditional CCTV systems rely on continuous human monitoring, which is inefficient and error-prone. Face recognition provides a reliable and nonintrusive biometric solution for real-time identification. It is highly suitable for smart surveillance applications due to its ease of deployment and accuracy.

This paper presents an IoT-based face recognition AI robot for intelligent surveillance. The system integrates ESP32-CAM for image capture, Python-based AI for recognition, and Arduino-controlled robotics for movement. It operates locally without cloud dependency to ensure low latency and improved privacy. The proposed solution is cost-effective, scalable, and suitable for smart environments.

## II. RELATED WORK

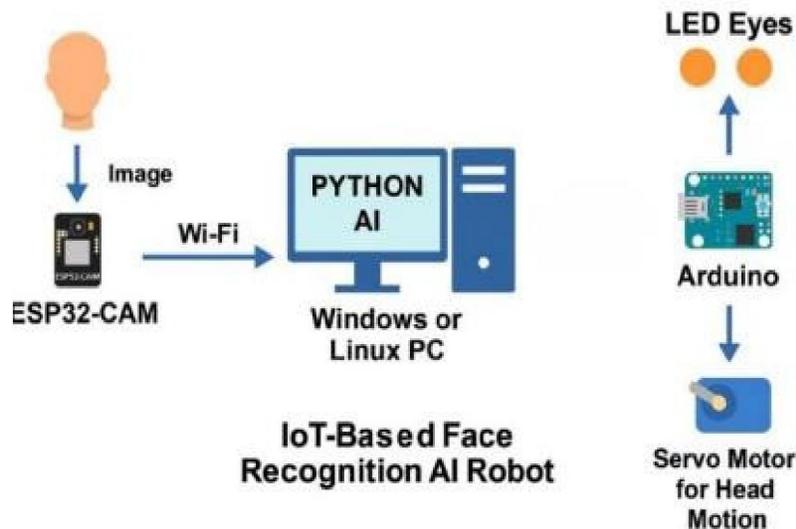
Face recognition research has evolved significantly over the past few decades. Early approaches relied on statistical and appearance-based techniques such as Eigenfaces, Fisherfaces, and Local Binary Patterns (LBP). Although computationally efficient, these methods suffered from limited robustness under varying illumination, pose, and facial expressions. The introduction of deep learning, particularly Convolutional Neural Networks (CNNs), revolutionized face recognition by achieving superior accuracy. Models such as FaceNet, VGG-Face, and DeepFace demonstrated near-human performance on benchmark datasets. However, these models require high computational resources, large memory, and often GPU acceleration, limiting their suitability for embedded and IoT-based systems. Recent studies have explored Raspberry Pi and ESP32-based face recognition systems for applications such as smart attendance, access control, and security monitoring. While these systems demonstrate feasibility, many rely on cloud-based processing, which introduces latency, bandwidth dependency, and privacy concerns. Additionally, most existing solutions lack integrated robotic interaction, limiting their ability to respond autonomously to detected events. In



contrast, the proposed system focuses on a fully local, embedded-friendly face recognition approach combined with robotic motion and interaction.

### III. SYSTEM ARCHITECTURE

The proposed IoT-based face recognition robot is designed using a modular and layered architecture to ensure scalability and efficient resource utilization. The system architecture consists of four primary layers: image acquisition, AI processing, robotic control, and communication.



The image acquisition layer employs the ESP32-CAM module, which captures real-time facial images and video streams. ESP32-CAM is selected due to its low cost, integrated Wi-Fi capability, and sufficient image quality for face recognition tasks. Captured frames are transmitted to the AI processing unit over a wireless connection.

The AI processing layer is implemented using Python and OpenCV. This layer performs face detection, feature extraction, and recognition. Facial landmark-based encoding is used to generate compact numerical representations of detected faces, which are stored locally for comparison.

The robotic control layer is handled by an Arduino microcontroller, responsible for controlling servo motors, LEDs, and sensors. This separation of processing and control tasks ensures that real-time motion control is not affected by computational delays in AI processing.

The communication layer enables data exchange between ESP32-CAM, the Python AI module, and the Arduino using Wi-Fi and serial/Bluetooth communication. This modular design allows independent upgrades and maintenance of system components.

### IV. FACE RECOGNITION METHODOLOGY

The face recognition process consists of two main phases: Enrollment and Recognition.

During the enrollment phase, facial images of authorized individuals are captured under controlled conditions. Each image undergoes preprocessing steps such as resizing, normalization, and grayscale conversion. Facial landmarks, including eye positions, nose shape, and jawline structure, are extracted using a landmark detection algorithm. These landmarks are encoded into a fixed-length feature vector representing the unique facial characteristics of the individual. In the recognition phase, live video frames are continuously captured and analyzed. Detected faces are encoded using the same landmark-based technique and compared with stored feature vectors using Euclidean distance. A predefined threshold value determines whether the detected face matches a known individual or is classified as unknown.



The face recognition process in the proposed system is designed to operate efficiently on embedded hardware while maintaining reliable recognition performance. Real-time image frames are captured using the ESP32- CAM module and transmitted to the processing unit for analysis. Each frame undergoes preprocessing steps such as resizing, grayscale conversion, and noise reduction to enhance image quality and reduce computational complexity. Face detection is performed using an efficient OpenCV-based detector, which identifies facial regions within the frame and isolates them for further processing.

After face detection, distinctive facial landmarks such as the eyes, nose, mouth, and jawline are extracted from the detected facial regions. These landmarks are used to generate a fixed-length numerical feature vector that represents the unique facial characteristics of an individual. A landmark-based encoding approach is employed to ensure robustness against minor variations in facial expression and pose while remaining computationally lightweight. During the enrollment phase, facial encodings of authorized individuals are generated and stored locally to form a reference database.

In the recognition phase, feature encodings obtained from live video frames are compared with the stored database using Euclidean distance metrics. A predefined threshold value is used to determine whether the detected face matches a known identity or is classified as unknown. Threshold optimization is performed experimentally to balance false acceptance and false rejection rates. Based on the recognition result, the system generates appropriate responses, such as audio feedback and robotic interaction for recognized individuals or alert mechanisms for unknown faces, enabling intelligent and autonomous system operation.

Threshold selection plays a crucial role in system performance. A lower threshold reduces false acceptance but may increase false rejection, while a higher threshold improves recognition flexibility at the risk of misidentification. Experimental evaluation is conducted to identify the optimal threshold value for embedded operation.

## **V. ROBOT INTERACTION AND CONTROL**

Human-robot interaction is a key feature of the proposed system. The Arduino microcontroller controls servo motors to enable head movement, allowing the robot to orient itself toward detected individuals. An ultrasonic sensor is used to detect human presence within a predefined range, triggering system activation.

The robot interaction and control subsystem is designed to enable responsive and intuitive human-robot interaction while maintaining real-time performance. An Arduino microcontroller serves as the primary control unit for handling actuators and sensors, allowing the system to operate independently of the AI processing module. This separation ensures that motion control and interaction tasks are not affected by computational delays associated with face recognition. Communication between the processing unit and the control unit is established through a wireless or serial interface, enabling coordinated system operation.

Servo motors are employed to provide controlled robotic movements, such as head rotation and orientation toward detected individuals. These movements enhance interaction by allowing the robot to track human presence and respond dynamically to changes in position. An ultrasonic sensor is used to detect proximity, activating the system when a person enters a predefined range. Visual feedback is provided through LEDs, which indicate system states such as detection, recognition, and alert conditions. This multimodal interaction improves system responsiveness and user awareness.

Audio interaction is implemented using a Python-based Text-to-Speech (TTS) module, enabling the robot to deliver verbal greetings or warning messages based on recognition outcomes. When a registered individual is identified, the system generates personalized responses, while unknown individuals trigger alert messages or logging mechanisms. The integration of motion, visual cues, and audio feedback allows the robot to function as an interactive surveillance agent rather than a passive monitoring device, thereby enhancing usability and effectiveness in real-world environments.

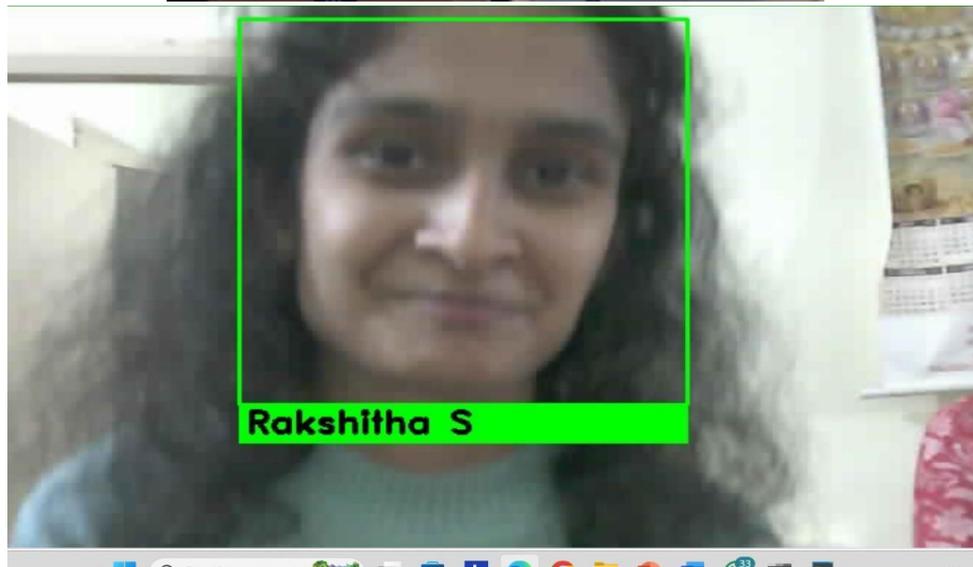
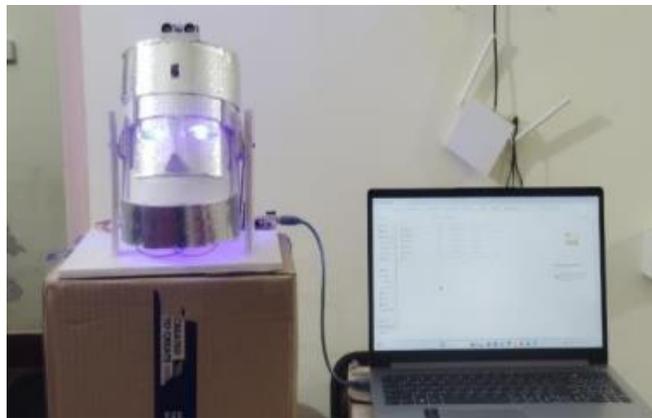


Visual feedback is provided using LEDs, which indicate system states such as detection, recognition, and error conditions. Audio feedback is implemented using Python-based Text-to-Speech (TTS), allowing the robot to greet recognized individuals or issue alerts when unknown faces are detected.

This multimodal interaction enhances user engagement and improves the effectiveness of surveillance by providing immediate and intuitive feedback.

### **VI. EXPERIMENTAL RESULTS AND DISCUSSION**

Experimental evaluation was conducted to assess the performance, reliability, and real-time capability of the proposed IoT-based face recognition AI robot. The system was tested in an indoor environment under varying lighting conditions and user distances to simulate practical surveillance scenarios. Multiple individuals were enrolled in the system database, and recognition experiments were performed using live video streams captured by the ESP32-CAM module. Key performance metrics such as recognition accuracy, response time, and system stability were analyzed to evaluate overall effectiveness.



The recognition accuracy of the system was measured by comparing correctly identified faces against the total number of test instances. Experimental results indicate that the system achieves recognition accuracy in the range of 90% to 100% under normal lighting conditions. Threshold analysis revealed that a moderate threshold value provides an



optimal balance between false acceptance and false rejection rates. As the threshold value increases, the system becomes more tolerant to variations but may incorrectly classify unknown faces, whereas lower thresholds improve security at the cost of increased false rejections. The selected threshold ensured stable recognition performance while maintaining system robustness.

Response time analysis demonstrated that the system is capable of real-time operation on embedded hardware. The average time from face detection to recognition output was observed to be less than two seconds, while servo motor actuation and audio feedback were triggered with minimal delay. The system operated continuously for extended durations without failure, indicating stable performance. Although slight degradation in accuracy was observed under low-light conditions and partial occlusions, the overall results validate the effectiveness of the proposed approach for real-time intelligent surveillance. These findings confirm that reliable face recognition and interactive robotic behavior can be achieved using low-cost embedded platforms without reliance on computationally intensive models.

### VII. APPLICATIONS

The proposed IoT-based face recognition robot can be deployed in various real-world scenarios, including smart surveillance systems, attendance monitoring in educational institutions, smart homes, and assistive robotics. Its low cost and modular design make it suitable for both academic and commercial applications.

- **Attendance Monitoring** The robot can automate attendance recording in educational institutions and workplaces by recognizing registered individuals as they enter a room. This reduces manual effort and prevents proxy attendance.
- **Smart Home Automation** In smart home environments, the system can recognize residents and trigger personalized actions such as unlocking doors, adjusting lighting, or activating security modes through IoT integration.
- **Human–Robot Interaction (HRI)** The integration of face recognition with speech and motion enables natural human–robot interaction. The robot can act as a receptionist, guide, or interactive assistant by recognizing familiar users and responding accordingly.

### VIII. CONCLUSION

This paper presented the design and implementation of an IoT-based face recognition AI robot developed for intelligent surveillance and human–robot interaction applications. The proposed system effectively integrates ESP32-CAM for real-time image acquisition, Python-based artificial intelligence for facial detection and recognition, and Arduino-controlled robotic mechanisms for motion and interaction. By employing a facial landmark-based encoding technique, the system achieves reliable recognition performance while maintaining low computational complexity, making it suitable for deployment on resource-constrained embedded platforms.

In conclusion, the integration of AI, IoT, and robotics in the proposed face recognition robot highlights the practicality of intelligent embedded surveillance systems. The work contributes a reliable and extensible framework that can be further enhanced through advanced algorithms, improved sensing capabilities, and secure IoT communication, paving the way for future developments in smart surveillance and autonomous robotic systems.

### IX. FUTURE WORK

Although the proposed IoT-based face recognition AI robot demonstrates reliable real-time performance using low-cost embedded hardware, several enhancements can be explored to further improve its robustness, scalability, and functionality. Future work will primarily focus on extending the system’s capabilities to address existing limitations and adapt it for large-scale real-world deployments.

Another promising direction for future work is the development of advanced human–robot interaction capabilities. Incorporating natural language processing (NLP), emotion recognition, and gesture recognition can enable more intuitive and context-aware interactions. This would allow the robot to respond dynamically to user behavior and environmental cues.



Finally, future research can explore multi-robot coordination and autonomous navigation using simultaneous localization and mapping (SLAM) techniques. This would enable cooperative surveillance using multiple robotic units, increasing coverage and reliability in large or complex environments. These enhancements would further establish the proposed system as a comprehensive and intelligent surveillance solution.

#### REFERENCES

- [1]. Jang, H., & Kim, Y. (2021). Development of an IoT-Based Smart Surveillance Robot with Face Recognition Function. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 12(7), 458–465.
- [2]. Dey, A., & Roy, S. (2021). A Study on Artificial Intelligence-Based Human–Robot Interaction Using IoT. *Journal of Intelligent Systems*, 30(1), 563–574.
- [3]. Verma, P., & Singh, R. (2021). Design and Development of a Low-Cost Face Recognition Robot Using Embedded Systems. *International Journal of Robotics and Automation Engineering*, 8(2), 33–40.
- [4]. Ahmed, S., & Rahman, M. (2022). Integration of AI and IoT in Robotics for Smart Interaction. *IEEE Access*, 10, 75560–75572.
- [5]. HC-SR04 Ultrasonic Sensor Datasheet. (2023). Distance Measurement and Obstacle Detection Module. Retrieved from <https://components101.com/sensors/hc-sr04-ultrasonic-sensor>

