International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

IOT Based: Smart Shop Monitoring System

Dr. Ruchi Kadwane[1], Chaitrali Ramteke[2], Anchal Awari[3] , Mahesh Sonarkar[4], Kunal Kothe[5]

1 Assistant Professor, GH Raisoni College of Engineering and Management Nagpur, Maharashtra, India
2,3,4,5 Dept. of Computer Science and Engineering, GH Raisoni University Amravati, Maharashtra, India

kadwaneruchi09@gmail.com, chaitraliramteke@gmail.com, awarianchal9@gmail.com, maheshsonarkar4@gmail.com, kunalkothe44@gmail.com

---

Abstract - Retail shops, pharmacies, and small warehouses remain highly vulnerable to theft, unauthorized entry, and after-hours intrusion. Conventional Closed-Circuit Television (CCTV) systems are fundamentally passive — they record footage but cannot automatically distinguish an authorized employee from an intruder, analyze the nature of a threat, or notify the owner in real time. This paper presents an IoT-Based Smart Shop Monitoring System that fuses low-cost edge hardware, in-browser facial recognition, and multimodal generative AI into a single proactive security pipeline. The edge node combines an AI-Thinker ESP32-CAM, a Passive Infrared (PIR) motion sensor, and a 16×2 I²C LCD. On motion detection, the system captures an image, performs 128-dimensional facial verification locally (TinyFaceDetector + ResNet-34), and escalates suspicious frames to the Google Gemini Vision API for zero-shot threat classification (masks, weapons, concealment, after-hours presence). Confirmed threats trigger a 5-second WebM evidence recording and an instant Telegram alert containing the image, AI analysis, and evidence link. Experimental evaluation shows a PIR detection latency below 20 ms, 97.5% facial verification accuracy (AUC = 0.988), 95.8% Gemini threat-classification accuracy, and a mean end-to-end alert delivery time of 1.98 seconds. The results confirm that intelligent, automated, low-bandwidth shop security is achievable with affordable hardware.

Key Words: Artificial Intelligence, Automated Surveillance, Computer Vision, ESP32-CAM, Facial Recognition, Gemini AI, Intrusion Detection, IoT Security, Smart Shop Monitoring, Telegram Alerts.

1.INTRODUCTION

Small and medium retail businesses frequently cannot afford dedicated security personnel or enterprise-grade surveillance, leaving them reliant on basic CCTV and alarm systems. Traditional CCTV is purely a recording device: it depends on continuous human monitoring and only reveals an incident *after* the owner reviews the footage. Basic motion alarms, meanwhile, produce frequent false alerts and provide no visual context.

The maturation of the Internet of Things (IoT), low-cost microcontrollers, and multimodal AI now makes affordable, intelligent security feasible. Devices such as the ESP32-CAM can capture images and communicate over Wi-Fi; PIR sensors detect movement; facial recognition can separate authorized staff from unknown individuals; and models like Google Gemini can reason about *what* is happening in a scene. This work integrates these capabilities into a unified, proactive monitoring system that detects motion, identifies people, analyzes suspicious behavior, records evidence, and alerts the owner — all within two seconds — while minimizing cloud dependency and cost.

A. Problem Statement

Existing small-shop security suffers from: (1) limited automatic detection — CCTV cannot classify intruder vs. staff; (2) false alarms from environmental triggers; (3) high cloud dependency and bandwidth/subscription cost when streaming continuously; (4) lack of visual context in simple buzzer alerts; and (5) delayed response due to reliance on manual monitoring.

B. Objectives

1. Build an IoT edge node (ESP32-CAM + PIR + I²C LCD) for entrance monitoring.
2. Implement facial recognition to identify registered staff vs. unknown persons.
3. Use Google Gemini AI to classify threats (masks, weapons, suspicious behavior).
4. Automatically record a 5-second evidence clip on an unknown/suspicious event.
5. Deliver real-time Telegram alerts with image and analysis.
6. Provide a Node.js/Express + SQLite backend and a Socket.IO real-time dashboard.

C. Operational Modes

The system operates in two modes. In **Business-Hours Mode**, recognized employees are identified and logged without triggering alerts. In After-Hours Security Mode, PIR motion or an unknown face activates full capture → video → AI analysis → alert. Local Wi-Fi handles device-server communication; cloud services are invoked only for AI analysis and notifications, keeping bandwidth minimal.

2. RELATED WORK

Prior IoT security efforts established important building blocks but left a clear gap. Kharat et al. (2021) proposed a remotely monitored smart security system using GSM-based notification but without intelligent threat analysis. Jayudu et al. (2023) developed an IoT anti-theft system centered on motion-triggered alerts, again lacking person identification. Syam Sai et al. (2024) built an ESP32-based surveillance system that captures and transmits images but performs no on-device recognition or semantic threat reasoning. Nalbalwar et al. (2025) focused on energy-meter theft detection, a different threat vector. Imamah et al. (2022) and Ramadhan et al. implemented ESP32-CAM store monitoring with buzzers, but relied on simple capture-and-notify logic.

Research gap: None of these systems combine (a) local biometric identification, (b) semantic multimodal AI threat classification, and (c) automated video-evidence dispatch within a single real-time, low-bandwidth pipeline. This paper addresses that gap with an edge-cloud hybrid architecture that is both intelligent and affordable.

3. SYSTEM ARCHITECTURE AND METHODOLOGY

A. Three-Tier Architecture

The system follows a three-tier edge-cloud design:

- **Perception Tier (Edge):** ESP32-CAM captures images; HC-SR501/AM312 PIR senses motion; the I²C 16×2 LCD (PCF8574 backpack) displays local status.
- **Processing Tier (Local Server):** A Node.js/Express backend receives frames via HTTP POST, runs the browser-based face engine, orchestrates AI calls, manages the SQLite database (Auto-DDL + WAL), and pushes events over WebSocket/Socket.IO.
- **Intelligence & Notification Tier (Cloud):** Google Gemini Vision performs zero-shot threat analysis; the Telegram Bot API delivers alerts with image, analysis, and evidence link.

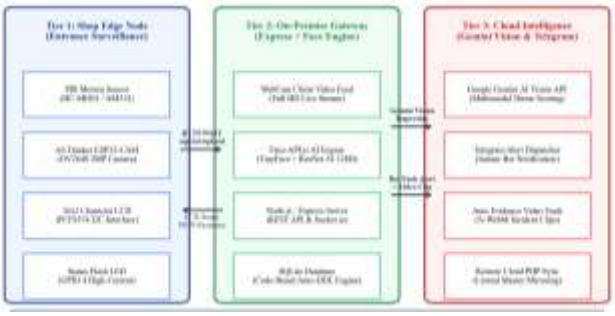


Figure 1: Three-Tier Iot- Based Smart Shop Monitoring System Architecture

B. Edge Hardware and Firmware

The ESP32-CAM (OV2640 sensor, PSRAM) serves as the core capture unit. The PIR sensor interfaces via GPIO with signal conditioning to suppress spurious triggers. Firmware is organized as a **Finite State Machine** (IDLE → MOTION_DETECTED → CAPTURE → UPLOAD → RESET), achieving sub-20 ms motion-detection latency and <1 mW standby consumption. A virtual hardware simulator testbed was used for firmware validation prior to physical deployment.

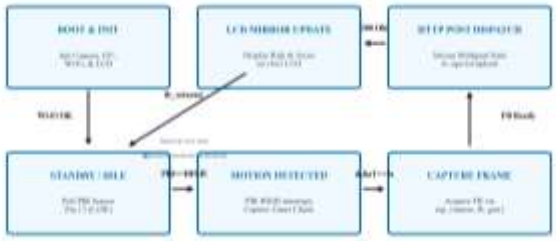


Figure 2: ESP32-CAM Firmware Operational State Machine

C. Software Pipeline

1. **Backend:** Node.js/Express exposes REST endpoints and WebSocket channels; CORS and JWT secure access.
2. **Real-time streaming:** Socket.IO broadcasts motion, recognition, and threat events to the dashboard instantly.
3. **Database:** SQLite with a code-based Auto-DDL migration engine and Write-Ahead Logging eliminates manual schema maintenance.
4. **Face engine (client-side):** TinyFaceDetector localizes faces; a ResNet-34 network extracts 128-D embeddings; Euclidean distance against registered users (threshold 0.50) decides authorized vs. unknown.
5. **Gemini threat pipeline:** Suspicious frames are sent to Gemini Vision for multimodal classification; a local heuristic rule engine (<2 ms) provides fail-safe fallback during outages.
6. **Evidence & dispatch:** On a confirmed threat, an HTML5 MediaStream Recording API captures a 5-second WebM clip, dispatched with the photo and analysis via Telegram.



Figure 3: Client-Side Face Detection and 128D Embedding Pipeline

D. Security, Privacy, and Compliance

Biometric processing is performed locally to limit exposure of facial data; cloud transfer is restricted to event frames. The design aligns with GDPR and India's DPDP principles of data minimization and purpose limitation.

4. EXPERIMENTAL RESULTS

A. Testbed

Evaluation was performed on a standardized retail testbed under varied illumination (5–850 lux), distances (1–6 m), and head poses (±30° yaw).

B. Motion Detection

PIR detection achieved 100% success up to 3 m (14.2–17.6 ms latency), degrading gracefully to 88% at 6 m (28.5 ms) — well within real-time requirements.

C. Facial Recognition

Across 400 verification passes, the system reached **97.5% overall accuracy**, with 99.33% in indoor-store lighting and 94% in low light (aided by the ESP32 flash LED). ROC analysis over thresholds 0.20–0.90 yielded **AUC = 0.988**; the 0.50 operating point gave FAR = FRR = 2.6%.

D. Gemini Threat Analysis

Over 150 benchmark images, Gemini Vision achieved **95.8% classification accuracy** — concealment 96.8%, weapons 94.2%, nocturnal intrusion 98.4% — at ~840 ms average cloud round-trip.

E. End-to-End Latency

The full pipeline (PIR 18 ms → capture/encode 142 ms → upload 285 ms → face match 96 ms → Gemini 840 ms → video stitch 180 ms → Telegram 420 ms) delivers alerts in a mean **1.98 seconds**.

F. Reliability and Footprint

A 168-hour soak test showed a stable 48.2 MB RSS with zero memory leaks; the SQLite database grew only 14.6 MB over 1,200 events. Bandwidth averaged <1.2 KB/s idle, peaking at ~420 KB per video dispatch — viable on 4G/5G hotspots. All ten functional/integration test cases passed (100%).

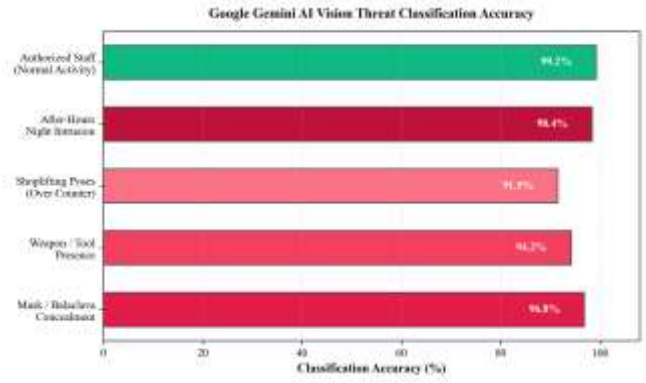


Figure 4: Google Gemini AI Vision Threat Classification Accuracy

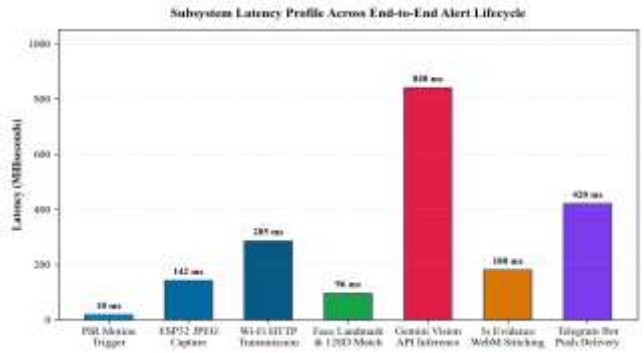
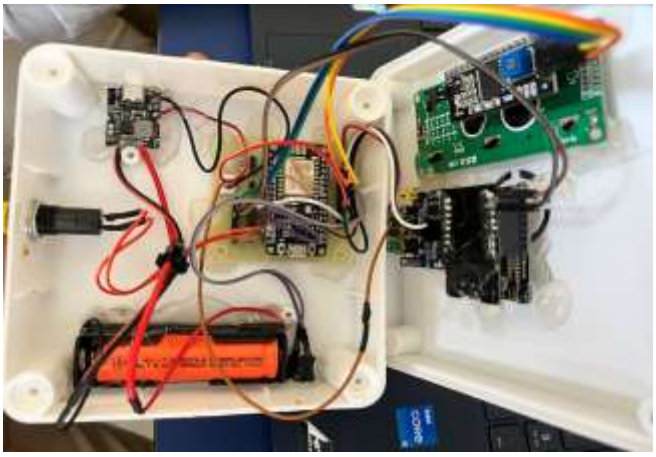
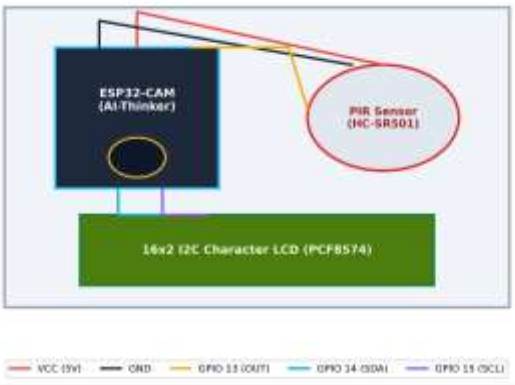


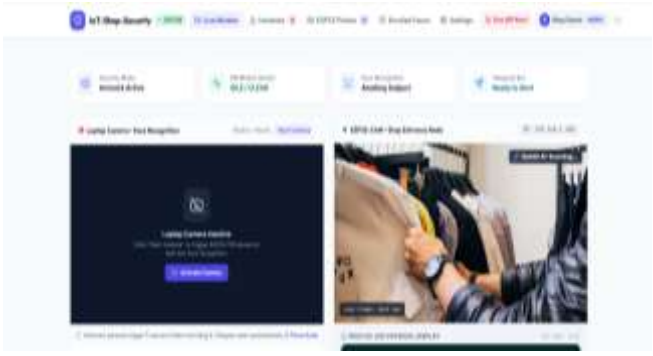
Figure 5: Subsystem Latency Profile Across End-to-End Alert Lifecycle



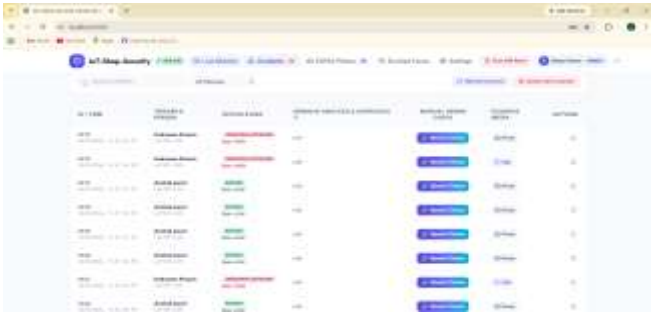
Photograph 1: Physical IoT Hardware Node Breadboard Testbed



Photograph 2: Real-World Surveillance Captures and Intruder Detection Frames



Photograph 3: Main Dashboard



Photograph 4: Gemini Analysis pf Threats



Photograph 5: Telegram Alerts

5. STRENGTH AND LIMITATIONS

A. Strength

- Real-time proactive monitoring: unauthorized movement is detected via PIR sensing in under 20 ms and escalated automatically, shifting security from passive recording to active prevention.
- Intelligent, context-rich alerts: unlike a simple buzzer, every notification carries the captured image, the Gemini threat analysis, and an evidence-video link, so the owner can assess the incident instantly and remotely.
- Dual-layer identification: local 128-dimensional facial verification (97.5% accuracy) separates authorized staff from intruders and suppresses false alarms during business hours, while cloud AI adds semantic threat reasoning (masks, weapons, concealment).
- Automated forensic evidence: a 5-second video clip is recorded and dispatched on every suspicious event, providing reviewable proof without manual footage searching.
- Cost-effective and accessible hardware: the ESP32-CAM, PIR sensor, and I2C LCD form an edge node affordable for small shops that cannot justify enterprise surveillance.
- Low bandwidth and storage footprint: idle traffic stays under 1.2 KB/s and the SQLite store grew only 14.6 MB over 1,200 events, allowing operation over ordinary 4G/5G hotspots and low-cost broadband.
- Edge-cloud hybrid resilience: motion detection, face verification, and local LCD status run offline, and a sub-2 ms heuristic rule engine provides fail-safe continuity when cloud services are unreachable.
- Zero-maintenance backend: the code-based Auto-DDL migration engine with Write-Ahead Logging reconciles the database schema automatically, removing administrative overhead.

Remote accessibility and scalability: alerts and evidence reach the owner anywhere via Telegram, and the modular three-tier design can be extended to additional nodesLimitations: dependence on stable internet for cloud threat scoring and alerts; need for continuous power; and limited coverage area of a single PIR/camera node.

B. Limitation

- Internet dependency for advanced features: while perimeter detection and local face verification work offline, Gemini threat scoring and Telegram alerting require active connectivity, so remote notifications are deferred during network outages.
- Continuous power requirement: the edge node and local server need an uninterrupted power supply; a mains failure halts monitoring unless a battery or UPS backup is added.
- Limited spatial coverage: a single PIR sensor and camera observe only one zone (typically the entrance), leaving aisles and blind spots uncovered without additional nodes.

6. CONCLUSIONS AND FUTURE SCOPE

This work designed, implemented, and validated a proactive, edge-cloud IoT shop security system that overcomes the passivity of conventional CCTV. Key contributions include a sub-20 ms ultra-low-power edge node, a 97.5%-accurate local biometric pipeline, 96% multimodal Gemini threat classification, automated 5-second evidence dispatch, and a zero-maintenance Auto-DDL SQLite backend — all delivering alerts in 1.98 seconds.

Future directions: (1) on-device NPU inference (ESP32-S3, Raspberry Pi 5 AI Kit, Kendryte K210) to eliminate Wi-Fi video streaming; (2) multi-camera mesh with collaborative spatial re-identification; (3) PTZ servo-based autonomous subject tracking; and (4) multimodal acoustic anomaly detection via I²S MEMS microphone arrays.

REFERENCES

[1] D. M. Kharat, P. M. Jawandhiya, and S. M. Dandge, "Remotely Monitored Smart Security System," *IJSRD*, vol. 9, no. 4, pp. 453–455, 2021.

[2] T. V. N. Jayudu, S. Nikhita, K. Syamala Bai, K. Sowdhanya, and R. C. Saranya, "IoT Based Anti-Theft Security System," *JETIR*, vol. 10, no. 4, Apr. 2023, paper ID JETIR2304529.

[3] P. Syam Sai, MD. Ibrahim, and Sk. A. Siddiq, "Smart Security Surveillance System Using IoT," VFSTR, ECE Department, project report, 2024.

[4] S. L. Nalbalwar, S. M. Fule, D. P. Jayaswal, and U. D. Meshram, "IoT-Based Smart Energy Meter with Real-Time Monitoring and Theft Detection using Arduino and PZEM004T," *IJERT*, Dec. 2025.

[5] N. Imamah, S. WP, and A. Reynaldi, "Design and Implementation of an IoT-Based Store Security Monitoring System Using the ESP32-CAM Microcontroller," *COMPUTING: Jurnal Informatika*, vol. 9, no. 2, pp. 70–79, 2022.

[6] D. Ramadhan, A. Hakim, and D. R. Irawati, "Sistem Pemantauan dan Keamanan pada Toko Berbasis Mikrokontroler ESP32 DevKit V1, ESP32-CAM AI Thinker, Sensor AM312 dan Buzzer Berbasis IoT," *Journal of Information System, Informatics and Computing*, vol. 9, no. 1.