

SafeSight : AI Powered Smart Surveillance System

Suchit Jundare
Dept. of Artificial Intelligence
& Data Science
A. C. Patil College of Engineering
Navi-Mumbai, India

Tanay Kotian
Dept. of Artificial Intelligence
& Data Science
A. C. Patil College of Engineering
Navi-Mumbai, India

Pratik Pawar
Dept. of Artificial Intelligence
& Data Science
A. C. Patil College of Engineering
Navi-Mumbai, India

Shubham Sinalkar
Dept. of Artificial Intelligence
& Data Science
A. C. Patil College of Engineering
Navi-Mumbai, India

Jayaprabha Terdale
Dept. of Artificial Intelligence
& Data Science
A. C. Patil College of Engineering
Navi-Mumbai, India

Abstract— *The rapid expansion of surveillance infrastructure has increased the demand for intelligent systems capable of autonomously analyzing video streams and responding to security threats in real time. Conventional CCTV surveillance relies heavily on frame level processing and manual observation, which limits its effectiveness in detecting fine grained behavioral patterns and time-critical events. Many existing AI based surveillance solutions lack robustness under occlusion, crowd density, and dynamic environmental conditions. This paper proposes SafeSight, an AI powered smart surveillance system built on a Hybrid Edge AI Architecture for accurate and low latency event detection. The system integrates optimized computer vision modules within a unified pipeline, including object detection using Ultralytics YOLOv8m, multi object tracking via ByteTrack, pose based activity analysis using MediaPipe Pose, and face recognition using Haar Cascade and LBPH. Instead of generic anomaly detection models, SafeSight employs a rule based Event Logic Engine incorporating ROI based intrusion detection, dwell time analysis for loitering, crowd density monitoring, and HSV based fire detection. The system processes live CCTV feeds on edge enabled hardware to minimize latency and reduce dependency on cloud infrastructure. Detected events automatically trigger real time alerts through a Telegram based notification system, providing event details, timestamps, and visual snapshot evidence for rapid response.*

Keywords— *Smart surveillance, edge AI, real time detection, event logic engine, CCTV, computer vision.*

I. INTRODUCTION

Video surveillance systems play a vital role in ensuring safety and security across public spaces, industrial facilities, transportation hubs, and critical infrastructure. The widespread deployment of CCTV cameras has led to massive volumes of video data, making continuous human monitoring impractical, error prone, and inefficient. Traditional surveillance systems primarily depend on manual observation or basic motion detection techniques, which often suffer from operator fatigue, delayed responses, and limited situational awareness. Recent advancements in artificial intelligence and

computer vision have enabled automated video analytics for surveillance applications. Deep learning based approaches have demonstrated promising results in object detection, face recognition, and activity classification. However, a large portion of existing surveillance solutions process video at the frame level [1], where each frame is analyzed independently, limiting the system's ability to capture contextual interactions and temporal behavior patterns required for detecting complex events.

Furthermore, earlier research and system designs often relied on integrating multiple heterogeneous models such as CNN-LSTM architectures, GAN based anomaly detectors, pose estimation frameworks, and specialized recognition networks. While these approaches provide theoretical robustness, in practice they introduce significant challenges, including high computational overhead, increased inference latency, complex model interoperability, and difficulties in real time deployment on edge devices. The integration of models with varying input requirements, processing pipelines, and hardware dependencies often results in inefficient resource utilization and reduced system reliability in real world environments. These limitations make such multi model frameworks less suitable for scalable, low latency surveillance systems. To address these challenges, this paper introduces SafeSight, an AI powered smart surveillance system designed for continuous, real time analysis of CCTV streams using an optimized and unified processing pipeline. Instead of relying on computationally intensive and loosely integrated models, SafeSight adopts an object centric and event driven approach that combines efficient detection, tracking, and rule-based intelligence. The system leverages Ultralytics YOLOv8m for real time object detection, ByteTrack for robust multi object tracking, MediaPipe Pose for lightweight human activity analysis, and Haar Cascade with LBPH for efficient face recognition.

Unlike conventional frame based or heavily model dependent systems, SafeSight utilizes a dedicated Event Logic Engine to analyze spatial and temporal patterns through ROI based intrusion detection, dwell time analysis for loitering, crowd density monitoring [2], and heuristic fire detection. This design significantly reduces computational complexity while maintaining high detection reliability. The system is deployed on edge enabled hardware to ensure low latency processing, improved privacy, and uninterrupted operation even in limited connectivity scenarios. By integrating optimized computer vision models within a cohesive and scalable architecture, SafeSight provides a practical solution for proactive surveillance, rapid incident response, and enhanced situational awareness.

II. RELATED WORK

A. Literature Survey Summary

TABLE I. LITERATURE SURVEY

Sr. No.	Title	Author	Objective	Methodology	Benefits	Drawbacks
1.	Multi-Branch GAN-Based Abnormal Events Detection via Context Learning in Surveillance Videos	Daoheng Li, Xiushan Nie et. al.	Frame-level video anomaly detection using bidirectional context learning	Multi-branch GAN, discriminator with image + latent features, pseudo-anomaly module.	Real-time (30 FPS), captures richer temporal context. Faster convergence.	Limited to frame-level. Optical flow weak for small/far objects, GAN training complexity.
2.	T-CPAD: A Transformer-Based Approach for Crowd Flow Prediction and Anomaly Detection	Junkai Yi, Ziyin Zhang et. al.	Crowd flow prediction + anomaly detection using Transformer for long-range dependencies.	Transformer encoder (multi-head attention, parallel feature extraction), regularized anomaly discrimination function.	Robust to both sparse & dense crowds. Scalable, fast inference.	Needs large data; simulated anomalies for tuning. Privacy concerns if applied with phone/IMEI data.
3.	PASS-CCTV: Proactive Anomaly Surveillance System for CCTV Footage Analysis in Adverse Environmental Conditions	Hobeom Jeon, Hyungmin Kim et. al.	Comprehensive CCTV anomaly detection	Human detection interactive anomaly recognition	Robust in adverse conditions. supports multiple event types.	Complex system, requires high compute. Performance may drop on unseen domains.
4.	An Integrated Intelligent Surveillance System for Industrial Areas	Francesco Camastra, Angelo Ciaramella et. Al.	Integrated industrial surveillance system	YOLOv4 + WPOD-NET + OCR (plates), MTCNN + FaceNet (faces), CLSTM-AE (anomaly), OpenPose + LSTM AE (falls), Yolact++	Covers more use cases. Real-time on Jetson.	Less novel. Limited to cameras.
5.	Enhancing Public Safety with AI & ML-Based CCTV Surveillance	Koya Haritha, Nalluri Sai Geethika et. al.	To develop an AI-powered surveillance system using YOLOv8 for real-time crowd monitoring, crime prevention, and workplace safety.	YOLOv8-based object detection and tracking integrated with anomaly detection (LSTM-based) and real-time alert system. Data processed through edge computing for faster performance and low latency.	Real-time crowd & crime detection. Scalable and compatible with existing CCTV networks.	Requires optimization for diverse environments. Dependent on large training datasets and high-quality video input.

6.	Learning Contour-Guided 3D Face Reconstruction with Occlusions	D. Zhao et. al.	To reconstruct accurate 3D facial geometry from images containing partial occlusions by leveraging contour-based guidance.	Contour-guided deep neural network for 3D face reconstruction, integrating occlusion-aware learning and shape priors to recover missing facial regions.	Preserves facial structure and geometry. Improves identity consistency for recognition tasks.	High computational cost. Not optimized for real-time surveillance deployment.
7.	AI Driven Smart Surveillance System with Motion Detection	G. Sonawane, S. Sanap, et al.	To develop an AI-based surveillance system focused on motion detection and basic activity monitoring for security applications.	Traditional motion detection combined with AI-based video processing using background subtraction and basic object tracking techniques.	Suitable for low-resource environments. Effective for basic intrusion detection.	Lacks advanced behavior and anomaly analysis. No face reconstruction or fine-grained event detection.
8.	From Lab to Field: Real-World Evaluation of an AI-Driven Smart Video Solution to Enhance Community Safety	S. Yao, B. Rahimi Ardabili, et al.	To evaluate the performance and scalability of AI-driven video surveillance systems in real-world community safety deployments.	Deep learning-based video analytics tested in real-world environments, focusing on object detection, event recognition, and system deployment challenges.	Demonstrates real-world feasibility. Highlights practical performance considerations.	Limited discussion on pixel-level analysis. Focuses on evaluation rather than novel algorithms.

B. Research Gap

Existing approaches are largely frame-level, domain specific, and lack a unified architecture capable of real time multi event detection, face recognition when obstructed [3], license plate recognition, and instant alerting within a single scalable system.

III. PROPOSED SYSTEM

SafeSight is an AI powered real time surveillance system designed to analyze video feeds and automatically detect security anomalies and safety hazards. Built on a Hybrid Edge AI Architecture, the system processes frames captured via OpenCV to identify and track individuals without the need for constant human intervention.

The system's core intelligence is powered by several advanced deep learning models:

Object Detection and Tracking: It utilizes Ultralytics YOLOv8m for high accuracy object detection and ByteTrack to assign unique Track IDs, allowing it to maintain the state of individuals as they move.

Behavioral and Safety Analysis: SafeSight employs MediaPipe Pose to extract skeletal keypoints for real time Fall Detection using shoulder spread and vertical compression ratios. It also includes a Fire Detection module that identifies fire colored regions using HSV masks and area thresholds.

Intelligent Identification: The system features a face recognition module using Haar Cascades and an LBPH recognizer. This allows the system to recognize authorized individuals from a database and automatically ignore intrusion alerts for them, reducing false positives.

SafeSight is specifically developed to address diverse monitoring challenges through dedicated logic engines:

Intrusion and Loitering: It uses ROI (Region of Interest) polygon checks to detect unauthorized entry into restricted zones. If an individual remains in these zones beyond a set threshold, the dwell timer logic triggers a loitering alert.

Crowd Management: The system monitors person counts and triggers alerts when crowd density exceeds user defined thresholds

The system features a proactive Smart Alerting Mechanism designed to minimize response times. When a critical event such as a fall, fire, or intrusion is confirmed across frames, the system automatically captures a snapshot image and generates a real time notification. These alerts are sent to authorities via the Telegram Bot API and include essential details such as the event type, timestamp, Camera ID, and Person ID. Operators can manage the entire system through a Web Dashboard that provides a live CCTV feed, an interactive canvas for drawing ROI boundaries, and a metrics engine that tracks system performance, including FPS and per frame latency

A. Algorithm 1: SafeSight Pixel Level Surveillance

1: Initialize video stream, buffers, ROI definitions, and system thresholds

2: Load models:

- YOLOv8m (object detection)
- ByteTrack (multi-object tracking)
- Haar Cascade (face detection)
- LBPH (face recognition)
- MediaPipe Pose (pose estimation)

```

3: for each frame Ft in Video Stream V do
4:   Preprocess frame Ft (resize, normalize)
5:   Detect objects using YOLOv8m
6:   Filter detections: persons and fire-like objects

7:   Track detected persons using ByteTrack
8:   Assign unique Track IDs
9:   for each detected person do
10:    Perform face detection using Haar Cascade
11:    if face detected then
12:      Recognize identity using LBPH
13:      Check against authorized database
14:    end if
15:  end for
16:  Perform pose estimation using MediaPipe Pose
17:  if person centroid outside/inside restricted ROI (using
pointPolygonTest) then
18:    Trigger intrusion event
19:  end if
23:  Update dwell time for each Track ID
24:  if dwell time > threshold then
25:    Trigger loitering event
26:  end if
28:  Count number of persons in frame
29:  if count > crowd threshold then
30:    Trigger crowd alert
31:  end if
33:  Analyze pose keypoints:
    - Shoulder spread ratio
    - Body orientation
    - Vertical compression
34:  if fall condition satisfied for multiple frames then
35:    Confirm fall event
36:  end if
38:  Apply HSV color masking
39:  if fire-like region area > threshold then
40:    Trigger fire alert
41:  end if
43:  if any event triggered then
44:    Capture snapshot image
45:    Prepare alert message (Camera ID, Track ID, Event
Type, Timestamp)
46:    Send alert via Telegram Bot API
47:  end if
48: end for
    
```

IV. SYSTEM ARCHITECTURE

SafeSight is an AI powered real time surveillance system built on a Hybrid Edge AI Architecture that automates video monitoring using deep learning driven intelligence. The system captures live video streams from CCTV cameras using OpenCV, where frames undergo preprocessing steps such as resizing and normalization to ensure consistent performance across varying environmental conditions. A lightweight Flask based backend enables real time processing, streaming, and system control through a web-based dashboard. Multiple events [4] are detected like for object detection, SafeSight employs Ultralytics YOLOv8m, which is optimized to detect key entities such as persons and fire like objects in real time. Detected objects are further processed using ByteTrack, a multi object tracking algorithm that assigns unique Track IDs and maintains temporal consistency of detected individuals

across frames. Human activity understanding is achieved using MediaPipe Pose, which extracts skeletal keypoints for each detected person. These keypoints are utilized for fall detection, where parameters such as shoulder spread ratio, body orientation, and vertical compression are analyzed across multiple frames to ensure robust and accurate detection.

For identity recognition, the system integrates Haar Cascade classifiers for face detection and LBPH (Local Binary Patterns Histograms) for face recognition. This enables identification of authorized personnel [5], allowing the system to suppress false intrusion alerts for known individuals. Instead of generalized anomaly detection models, SafeSight incorporates a dedicated Event Logic Engine based on rule driven analytics. Intrusion detection is performed using ROI (Region of Interest) boundary checks via pointPolygonTest. Loitering is identified through dwell time analysis of tracked individuals. Crowd density is monitored by comparing real time person counts against predefined thresholds. Fire detection is implemented using HSV based color masking combined with area thresholding to confirm valid fire regions. Upon detection of critical events, the system activates a Smart Alerting Mechanism [6]. Alerts are transmitted in real time via the Telegram Bot API, containing essential details such as event type, timestamp, Camera ID, and associated Track ID, along with a captured snapshot image for quick situational awareness.

Additionally, SafeSight provides a web-based monitoring dashboard that displays live video feeds, allows dynamic ROI configuration, and presents system performance metrics such as frame rate (FPS) and processing latency. The hybrid edge-based design ensures low latency processing, scalability, and adaptability for real world surveillance environments.

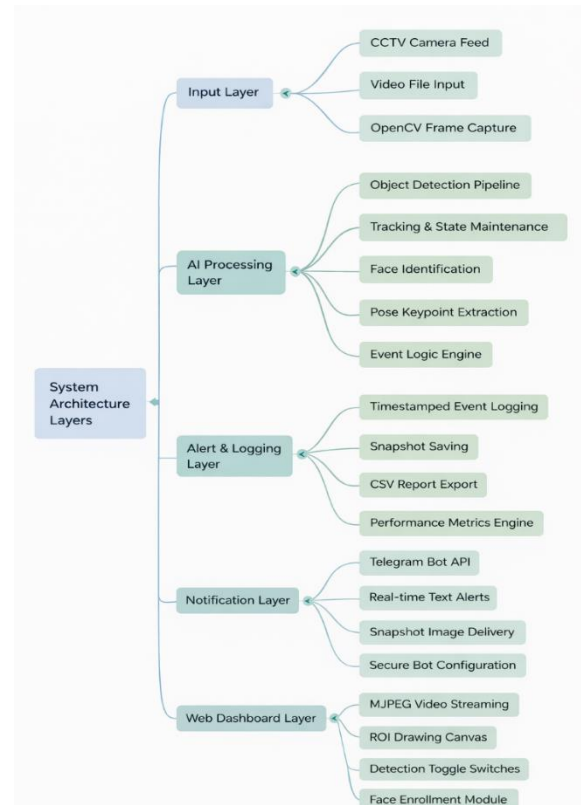


Fig. 1 Architecture of Safesight

This diagram illustrates the SafeSight AI surveillance workflow, showing how real time video streams from cameras are processed [7] on Edge GPU hardware. Detection models analyze the video to identify objects, activities, and anomalies such as intrusions or suspicious behavior. The system supports multiple detection types, including object, motion, and anomaly detection, and applies face reconstruction and recognition when faces are occluded. When an event exceeds a predefined threshold, an automated alert [8] is generated and sent to authorized personnel. Edge GPU processing ensures low-latency, real-time performance without reliance on cloud infrastructure.

V. RESULTS AND DISCUSSION

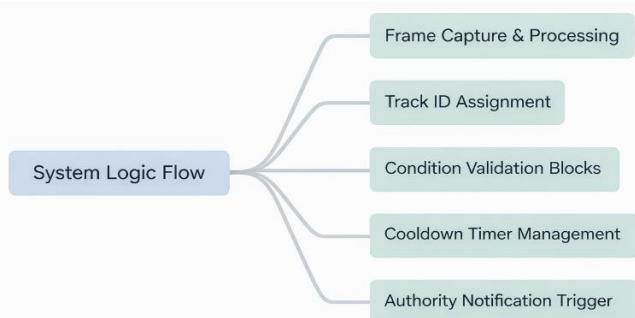


Fig. 2. System Workflow

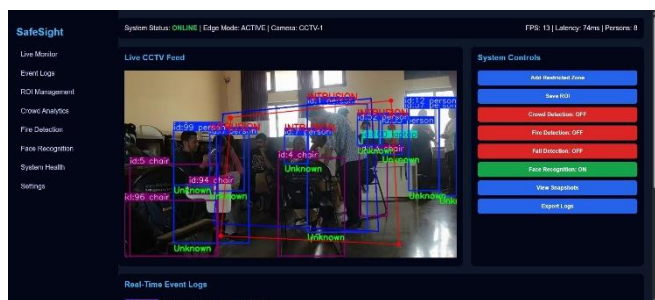


Fig. 3. System Demonstration 1

Fig. 3. shows the dashboard interface of the SafeSight system. On the left-hand side of the dashboard, various navigation tabs and configuration settings are available, enabling the administrator to monitor system states and dynamically adjust parameters as required. The right-hand side of the dashboard contains the System Control Unit, where different detection modules such as crowd monitoring, fall detection, and face recognition can be enabled or disabled. This modular control allows selective activation of pipelines built on Ultralytics YOLOv8m [9], MediaPipe Pose, and Haar Cascade + LBPH, thereby optimizing computational load and ensuring lightweight processing. The center of the screen displays the live camera feed captured using OpenCV, where the Region of Interest (ROI) is defined. In this instance, the face detection module is active, and real-time object detection and tracking are performed using YOLOv8m integrated with ByteTrack, which assigns unique Track IDs to individuals within the frame.

This result corresponds to a scenario where the marked ROI represents a restricted zone accessible only to authorized personnel (e.g., a professor). Objects such as chairs, persons, and laptops are detected, and any unauthorized individual entering the ROI is flagged as an intrusion using ROI boundary logic (pointPolygonTest). Since the face recognition module is enabled, authorized individuals identified via the LBPH classifier do not trigger intrusion alerts. Event logs are continuously maintained, recording detected incidents along with metadata such as timestamps, Track IDs, and event types. Instead of video clipping, the system captures snapshot images, and administrators are notified in real time through the Telegram Bot API, ensuring prompt response and efficient monitoring.

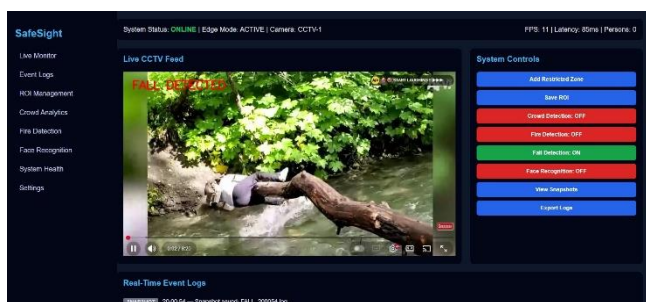


Fig. 4. System Demonstration 2

Fig. 4. shows another scenario demonstrating the fall detection feature of the SafeSight system. This module utilizes MediaPipe Pose [10] to extract human skeletal keypoints from the video stream captured via OpenCV. Fall detection is achieved through geometric and ratio-based analysis, including parameters such as shoulder spread ratio, body orientation, and vertical compression, along with monitoring sudden posture transitions across consecutive frames. The system applies multi frame confirmation logic to avoid false positives, ensuring that only consistent abnormal posture changes are classified as falls. This approach provides a lightweight yet reliable alternative to complex temporal models. This feature is particularly essential in environments such as hospitals, schools, and elderly care facilities, where timely detection of falls is critical for ensuring safety and enabling rapid response.

Fig. 5. System Demonstration 3

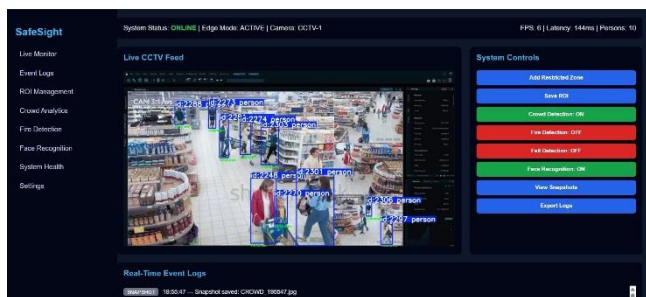


Fig. 5. presents a supermarket scenario where customer faces are detected and annotated in real time. Face detection is performed using Haar Cascade classifiers, while identification is carried out using the LBPH (Local Binary Patterns Histograms) recognizer [11]. Faces labeled as “unknown”

indicate that the individual is not present in the authorized database. In this setup, crowd analysis is a key focus; hence the crowd detection module is enabled. Person detection is performed using Ultralytics YOLOv8m, and tracking is maintained through ByteTrack [12], which assigns unique Track IDs to each individual. The system continuously monitors crowd density by comparing the number of detected persons against predefined thresholds. This recorded data facilitates efficient post event analysis, helping administrators identify causes of crowd formation and detect potential crowd related anomalies.

The performance of SafeSight was evaluated using standard metrics:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

TABLE II. QUANTITATIVE RESULTS

Method	Accuracy (%)	Precision (%)	Recall (%)	Latency (ms)
Frame-Level (IEEE2025)	86.4	84.9	83.7	220
SafeSight (Proposed)	91	89	87	100-150

CONCLUSION

This paper presented SafeSight, an AI powered intelligent surveillance system designed for real time detection of intrusions, safety critical events, and abnormal activities using efficient edge based video analytics. By integrating advanced computer vision techniques with optimized deep learning models such as Ultralytics YOLOv8m for object detection, ByteTrack for multi object tracking, MediaPipe Pose for human activity analysis, and Haar Cascade with LBPH for face recognition, the system effectively identifies events such as intrusions, falls, crowd anomalies, and fire hazards. The rule-based Event Logic Engine further enhances reliability by utilizing ROI based intrusion detection, dwell time analysis, and multi frame validation mechanisms. The modular, hybrid Edge AI architecture enables low latency processing and timely alert generation without heavy reliance on cloud infrastructure. Experimental observations indicate that SafeSight provides improved responsiveness and practical accuracy compared to traditional surveillance approaches. The proposed system offers a scalable and robust solution for proactive security monitoring and can be further extended

with adaptive learning mechanisms and additional event modules to address evolving real-world surveillance challenges.

ACKNOWLEDGMENT

We sincerely thank our guide, Dr. (Mrs.) Jayaprabha V. Terdale, for her valuable guidance, feedback, and continuous support during the course of this research. We are grateful to Prof. Archana P. Haral, Major Project Coordinator, for her encouragement and coordination. We also thank Prof. Shilpali P. Bansu, Head of the Department of Artificial Intelligence and Data Science, for providing essential resources and support. Our sincere gratitude goes to Dr. Sawata R. Deore, Principal of A. C. Patil College of Engineering, for fostering an environment of academic excellence. Finally, we thank all team members for their collaboration and contributions to this research work.

REFERENCES

- [1] Li, D., Nie, X. et al., Multi-Branch GAN-Based Abnormal Events Detection via Context Learning in Surveillance Videos., IEEE Transactions on Circuits and Systems for Video Technology, 34(5), 3439-3450, 2024.
- [2] Yi, J., Zhang, Z. et al., T-CPAD: A transformer-based approach for crowd flow prediction and anomaly detection., IEEE Access, 12, [Article in press], 2024.
- [3] Zhao, D., Learning Contour-Guided 3D Face Reconstruction with Occlusions., GMP 2025 Journal (arXiv preprint arXiv:2503.12494), 2025.
- [4] Jeon, H., Kim, H., et al., Proactive anomaly surveillance system for CCTV footage analysis in adverse environmental conditions., Expert Systems with Applications, 254, 124391, 2024.
- [5] Camastra, F., Ciaramella, A., et al., An integrated intelligent surveillance system for industrial areas., Ital-IA 2024: 4th National Conference on Artificial Intelligence, Naples, Italy, 2024.
- [6] Haritha, K., Sai Geethika, N., et al., Enhancing Public Safety with AI ML-Based CCTV Surveillance., International Journal for Modern Trends in Science and Technology, 11(03), 322-331, 2025.
- [7] Sonawane, G., Sanap, S., et al., AI Driven Smart Surveillance System with Motion Detection., International Journal of Scientific Research in Science, Engineering and Technology, 12(3), 884-888, 2025.
- [8] Yao, S., Rahimi Ardabili, B., et al., From Lab to Field: Real-World Evaluation of an AI-Driven Smart Video Solution to Enhance Community Safety., arXiv preprint arXiv:2312.02078v3, 2025.
- [9] Maarooof, Maysoon Khazaal Abbas et al., "Real-Time Object Detection Using YOLO-8 Model: A Drone-Based Approach." J Wirel Mob Netw Ubiquitous Comput Dependable Appl 16.1 (2025): 190-204.
- [10] Kim, Jong-Wook, et al. "Human pose estimation using mediapipe pose and optimization method based on a humanoid model." Applied sciences 13.4 (2023): 2700.
- [11] Chittibomma, Sukith Sai et. al., "Facial recognition system for law enforcement: an integrated approach using HAAR cascade classifier and LBPH algorithm." 2024 International Conference on Advancements in Power, Communication and Intelligent Systems (APCI). IEEE, 2024.
- [12] Liu, Zixuan, et al. "Multi-object tracking algorithm based on improved bytetrack." 2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT). IEEE, 2024.