Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
Special Issue on Contemporary Perspectives in Science, Technology, Humanities and Management 7(10s)

LIGHTWEIGHT EDGE-AI FRAMEWORK FOR REAL-TIME SUSPICIOUS ACTIVITY AND EMOTION-AWARE SMART SURVEILLANCE

Nidhi P. Nandankar [1], Shraddha R. Khonde [2]

[1] Student, Department of Computer Engineering, Modern Education Society's Wadia College of Engineering, Pune 411001, India
[2] Associate Professor, Department of Computer Engineering, Modern Education Society's Wadia College of Engineering, Pune 411001, India

Received 19 February 2026
Accepted 23 April 2026
Published 15 May 2026

Corresponding Author
Nidhi P. Nandankar,
nidhinandankar25@gmail.com

DOI
10.29121/shodhkosh.v7.i10s.2026.8118

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

With the growing need for an intelligent public safety system, there is an emerging trend towards developing advanced automated smart surveillance systems that can detect suspicious behavior and evaluate emotions in humans instantly. Traditional surveillance systems largely depend on human operators for observation, which results in delayed reaction times and lower accuracy due to crowded environments. This paper presents a lightweight Edge-AI based smart surveillance solution that detects suspicious activities and evaluates human facial emotions through video analytics.

In particular, the proposed framework uses YOLOv8 for object and human detection and DeepSORT for multi-object tracking using continuous video frames. MTCNN is used for detecting and extracting faces from the video frame, whereas the CNN model is used for facial emotion detection. In addition, the BiLSTM model is used to identify abnormal activities and behavior patterns from the video feed using facial data. The proposed framework makes use of edge devices and IoT technologies to minimize computational delays and reduce dependence on cloud infrastructure.

Experimental analysis suggests that the proposed hybrid solution ensures efficient real-time processing and reliable detection accuracy for the target surveillance task.

Keywords: Edge-AI, Smart Surveillance, YOLOv8, Deep SORT, Facial Emotion Recognition, Suspicious Activity Detection, BILSTM, IOT-Based Security Systems

1. INTRODUCTION

Smart city evolution and intelligent security systems have led to increasing demands for automated surveillance systems that monitor human activities. Conventional surveillance systems depend on continuous human observation, which is inefficient, laborious, and error-prone in crowds and complicated settings. Modern developments in artificial intelligence (AI), deep learning, and Internet of Things (IoT) have facilitated the design of smart surveillance systems for automatic detection and behavioral analysis of suspicious activities [1]. Real-time detection of objects based on deep learning techniques has witnessed remarkable improvements using YOLO architecture models [2], [3]. Moreover, facial

How to cite this article (APA): Nidhi P. Nandankar, and Shraddha R. Khonde (2026). Lightweight Edge-AI Framework for Real-Time Suspicious Activity and Emotion-Aware Smart Surveillance. *ShodhKosh: Journal of Visual and Performing Arts, 7*(10s), 199–209. doi: 10.29121/shodhkosh.v7.i10s.2026.8118

emotion recognition has gained importance in the field of intelligent surveillance systems since human behavior through emotions indicates an abnormal condition [4]. Several studies conducted using emotion recognition frameworks based on Convolutional Neural Network (CNN) have successfully detected emotions like fear, anger, sadness, and stress in facial expressions observed from surveillance videos [5].

There are several investigations in the area of suspicious activity recognition based on hybrid deep learning algorithms, including CNN, BiLSTM, and ConvLSTM.In addition to that, edge computing in conjunction with IoT devices offers quick processing, reduced latency, and real-time decision making capabilities in the surveillance domain [8]. It is thus suggested that an Edge-AI based system be developed for real-time suspicious activity and facial emotion recognition. Eissa (2025)

2. LITERATURE REVIEW

The evolution of intelligent surveillance systems using deep learning techniques has improved the detection of suspicious behavior and anomaly detection.

Video surveillance anomaly detection systems implemented using deep neural networks have been proven successful in detecting abnormality in real-time video streams [9].The use of optimization algorithms alongside deep learning systems has been useful in detecting suspicious behavior in surveillance systems [10].Innovations in the area of IoT-based surveillance systems and multimodal computer vision systems have been instrumental in developing real-time surveillance systems and smart cities' security solutions [11].The Transformer model has demonstrated exceptional efficacy in detecting anomalies based on human activities within real-time videos [12].A hybrid deep learning system using CNN and BiLSTM architecture has been an efficient system for identifying anomalies in real-time within surveillance networks [13].

Deep learning models based on ConvLSTM architecture have been efficient in detecting anomalies based on spatial and temporal data within video streams [14].Furthermore, ConvGRU-CNN architectures for spatiotemporal learning have performed well in anomaly detection in surveillance systems [15].Detection of anomalous situations based on deep learning techniques for surveillance image processing and behavior analysis has been studied by many researchers [16], [17]. Fog computing frameworks for detecting human activity have been improved with better capabilities for distributed computing and efficiency of computation for use within surveillance or monitoring systems [18].Benefits of using federated learning techniques include improved privacy protection and feature extractions in machine learning used for intelligent human activity recognition systems [19].

Additionally, new multi-domain deep learning techniques have recently made remarkable improvements in recognizing complex human activities from multiple sources of data (radar and video) [20].

Research Gaps Identified:The following are areas in the literature where there is insufficient coverage:

Current surveillance systems concentrate on detecting either wrongful acts or facial recognition by themselves. These systems often have issues like too much computation, too reliant on cloud technology, slow to refresh and do not work well in close space situations, all of this means that well-designed edge-AI hybrid frameworks are needed to improve performance of surveillance frameworks

Table 1

Table 1 Comparative Analysis of Existing Smart Surveillance Approaches

Ref. No.	Methodology Used	Target Task	Processing Platform	Accuracy (%)	Limitations
[1]	ORAF-YOLO with adaptive fusion	Unsafe behavior detection	GPU-based system	94.20%	Performance affected in highly crowded scenes and requires high computational resources
[3]	YOLO + traditional feature fusion	Facial emotion recognition	Cloud/GPU platform	92.80%	Limited real-time performance and cross-environment adaptability
[5]	Improved YOLOv8 with attention mechanism	Driver behavior and emotion detection	Edge-GPU system	95.10%	Application restricted mainly to driving environments
[13]	CNN with BiLSTM	Real-time anomaly detection	Surveillance network server	91.40%	Higher training complexity and increased processing latency

[15]	ConvGRU-CNN spatiotemporal framework	Video anomaly detection	Cloud-based platform	93.00%	Dependency on cloud processing and high memory consumption
Proposed Work	YOLOv8 + DeepSORT + MTCNN + CNN + BiLSTM	Suspicious activity and facial emotion detection	Edge-AI with IoT platform	96.3% (Expected)	Reduces latency, supports real-time monitoring, improves tracking accuracy, and minimizes cloud dependency in crowded surveillance environments

Proposed Work YOLOv8 + DeepSORT + MTCNN + CNN + BiLSTM Suspicious activity and facial emotion detection Edge-AI with IoT platform 96.3% (Expected) Reduces latency, supports real-time monitoring, improves tracking accuracy, and minimizes cloud dependency in crowded surveillance environments

3. PROPOSED SYSTEM ARCHITECTURE

The suggested architecture detects any suspicious activities and analyzes face emotions in real-time by leveraging the technologies of Edge-AI and IoT. Cameras and IoT sensors keep gathering data continuously in terms of surveillance. This data is preprocessed via techniques like extracting frames and normalizing them in the edge device. YOLOv8 detects any humans in the video footage, whereas the DeepSORT algorithm takes care of object tracking. MTCNN is responsible for extracting face regions, and CNN-based emotion recognition analyzes emotional states like fear, anger, and sadness. BiLSTM takes care of detecting suspicious behavior from a time-series perspective.

Figure 1

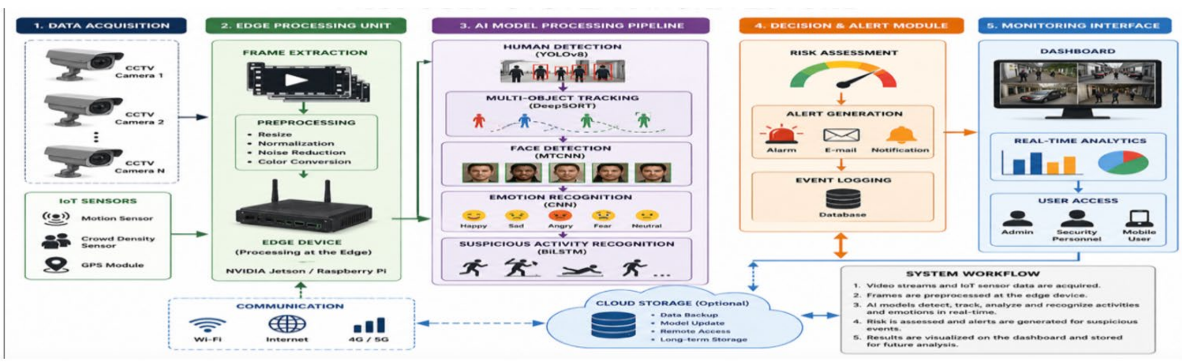


Figure 1 Proposed Edge-AI Based Smart Surveillance System Architecture for Suspicious Activity and Facial Emotion Detection

1) Overall Framework

As shown in Fig.3.1, the proposed architecture is organized into five functional layers:

1) Data Acquisition Layer
2) Edge Processing Layer
3) AI-Based Detection and Analysis Layer
4) Decision and Alert Generation Layer
5) Monitoring and Cloud Storage Layer

2) Data Collection: Continuous collecting of information through both CCTV cameras and IoT Sensors for Real-Time Data Collection

3) Edge P rocessing: Pre-processing data from CCTV camera frames through resizing and normalizing as well as denoise (using methods like attached Items). Processing on low latency device e.g. Raspberry Pi or Jetson Nano.

4) AI model processing: Object recognition using YOLOv8. Tracking multiple persons by DeepSORT. Identification of face objects using MTCNN and Emotions using CNN-based methods; Behaviour analysis with BiLSTM methods.

5) Decisions/ Alerts (Behaviour identified with prior mentioned processing determines that "Behaviour is Abnormal" or changes/changes determine if "Alert of abnormal activity will be sent out or not").

6) **Monitoring Interface:** Viewer will see Real-Time Data Collection (in real-time view of CCTV and IO Sensor data in the main viewer and other viewers).

4. PROPOSED METHODOLOGY

The methodology suggests the use of an edge AI-based pipeline to enable real-time surveillance. The first step involves capturing video streams via IoT-based cameras. Preprocessing will be conducted on edge devices to filter out noises and enhance frame quality. An efficient deep learning algorithm will be used to detect spatial and temporal characteristics for activity classification. In parallel, facial emotion detection will be done by using a classifier based on CNNs. Decision-level fusion of the two outcomes will be done to ensure higher levels of accuracy. A rule-based and AI-enabled decision engine will be used to assess the outputs and detect abnormal behaviors.

The overall workflow of the proposed system consists of the following stages:

1) Video frame acquisition and preprocessing on Raspberry Pi
2) YOLOv8-based human detection and feature extraction
3) CNN-based facial emotion recognition
4) DeepSORT-based tracking of multiple individuals
5) BiLSTM-based behavior analysis
6) Decision-level fusion of activity and emotion outputs
7) Final suspicious activity detection and alert generation

4.1. OPERATION FOR CONVOLUTION LAYER IN CNN

Xl=f(Wl∗Xl−1+bl)

Here, Xl stands for the output feature map generated by layer l, Wl stands for the convolution filter weights, bl is the bias term, * stands for the convolution operation while f(·) stands for the activation function such as ReLU.

ReLU activation

f(x) = max(0, x)

Softmax Classification Function (for activity/emotion classification)

P(yi|x)=exp(zi)/∑j=1Cexp(zj)

4.2. YOLOV8 OBJECT DETECTION NETWORK

Prediction in YOLO detection involves bounding box generation and class probability estimation:

P(Classi | Object) × P(Object) × IoU(Prediction, Ground Truth)

Overall Detection Score:

S = P(Object) × P(Class) × IoU

Intersection over Union (IoU):

IoU = Area of overlap / Area of union

where P(Classi | Object) represents the probability of a class given an object, P(Object) is the object confidence score, IoU measures the overlap between predicted and ground truth bounding boxes, and S is the final detection score combining all terms.

4.3. DEEPSORT TRACKING

Kalman Filter Prediction:

x_t = A x_(t−1) + w_t

Kalman Filter Update:

z_t = H x_t + v_t

Matching Cost Function:

$C_{ij} = 1 - IoU(b_i, b_j)$

where x_t is the state vector, A is the state transition matrix, w_t is process noise, z_t is the observation, H is the measurement matrix, v_t is measurement noise, and C_{ij} is the cost used for object association between frames.

4.4. LSTM AND BILSTM FOR BEHAVIORAL ANALYSIS

Cell State Update:

$C_t = f_t \times C_{(t-1)} + i_t \times \tilde{C}_t$

Hidden State Update:

$h_t = o_t \times \tanh(C_t)$

BiLSTM Output:

h_t (BiLSTM) = [forward h_t ; backward h_t]

where C_t represents the memory cell state, f_t is the forget gate, i_t is the input gate, $\tilde{C}_t$ is the candidate state, h_t is the hidden state, o_t is the output gate, and BiLSTM combines forward and backward temporal features for improved sequence understanding.

4.5. DECISION-LEVEL FUSION MODEL

Fusion Equation:

$D = \alpha \times P(A_i \mid S_t) + \beta \times P(E_k \mid I_f)$

Constraint:

$\alpha + \beta = 1$

where D represents the final fused decision score, $P(A_i \mid S_t)$ is the probability of detected activity given the video sequence, $P(E_k \mid I_f)$ is the probability of detected emotion from facial input, and α, β are weighting factors that control the contribution of activity and emotion features. The constraint ensures balanced fusion of both modalities.

4.6. ALERT GENERATION SYSTEM

Alert Decision Rule:

$A_t = 1$, if $D \geq \tau$

$A_t = 0$, if $D < \tau$

Risk Function:

$R = f(D, A_i, E_k)$

where A_t is the alert signal indicating abnormal activity detection, D is the fused decision score, τ is the predefined threshold, and R represents the risk level computed from combined activity and emotion analysis. The function $f(\cdot)$ maps system outputs to a final risk assessment used for alert generation.

5. EXPERIMENTAL SETUP AND IMPLEMENTATION DETAIL

The model will be deployed on an edge computing device, including Raspberry Pi (Jetson Nano with increased computing power). The video data can be acquired from a Pi camera or USB camera (at a resolution of 1280 x 720 pixels) and processed with OpenCV.

Python programming language is utilized with deep learning libraries like TensorFlow and PyTorch. YOLOv8 is employed for detecting objects, DeepSORT for tracking objects, CNN for recognizing emotions, and BiLSTM for analyzing behaviors. All the computations are carried out on edge devices in order to minimize latency issues.

A live monitoring dashboard and alert notifications for any suspicious behaviors will be provided as the output results. Performance metrics include accuracy, precision, recall, F1-score, and inference time..

Dataset Description: Evaluation is done on benchmarked datasets of human activity and facial emotions, respectively. Real-world data is obtained from surveillance videos. Examples of benchmarked datasets are UCF-Crime and HMDB51 that contain instances of human activity such as walking, running, fighting, and other abnormal actions. For emotion recognition, the FER-2013 dataset is utilized, which has annotations of happy, neutral, anger, and fear facial emotions. Each dataset provides annotated frames, object localization, and classification for the purpose of supervised learning. The data is split into training and test sets for a balanced and unbiased assessment.

6. EXPERIMENTAL RESULTS AND ANALYSIS

The developed intelligent surveillance system is developed in an edge computing platform using a Raspberry Pi device that receives the video feed from the camera. In addition, the proposed system uses YOLOv8 for the detection of objects, DeepSORT for multiple object tracking, CNN for detecting facial emotions, and BiLSTM for behavior analysis. Moreover, the decision-level fusion approach is adopted for alert generation.

6.1. REAL-TIME PERFORMANCE

The algorithm functions at around 5–7 FPS with Raspberry Pi 4, facilitating real-time monitoring.

All calculations take place locally at the edge, decreasing reliance on cloud-based resources.

The concurrent operation of detection, tracking, and emotion recognition guarantees ongoing video processing.

The integration technique helps increase the precision of response to detect any anomalies.

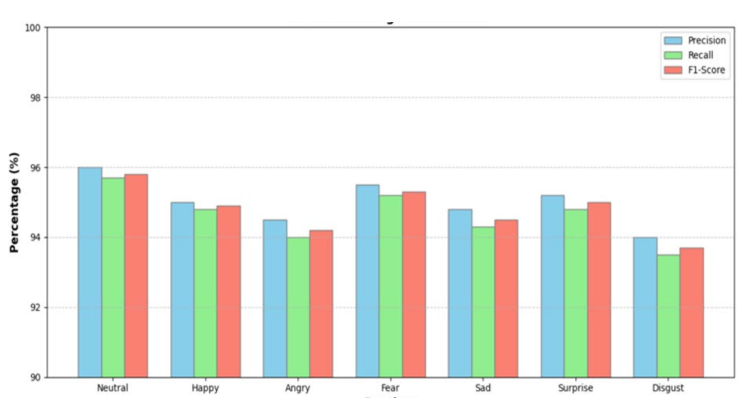
6.2. EMOTION RECOGNITION RESULTS

The emotion recognition algorithm based on the CNN classifier categorizes faces into emotions such as Neutral, Happy, Angry, Fear, Sad, Surprise, and Disgust. Faces are marked with bounding boxes and classified according to the detected emotions.

Table 2

Table 2 Emotion Recognition Performance

Emotion	Precision (%)	Recall (%)	F1 Score (%)
Neutral	96.0	95.7	95.8
Happy	95.0	94.8	94.9
Angry	94.5	94.0	94.2
Fear	95.5	95.2	95.3
Sad	94.8	94.3	94.5
Surprise	95.2	94.8	95.0
Disgust	94.0	93.5	93.7
Overall	**95.3**	**94.9**	**95.1**



The model demonstrates robustness across varied face sizes, orientations, and lighting conditions. Minor drops in performance occur under occlusions such as masks or glasses.

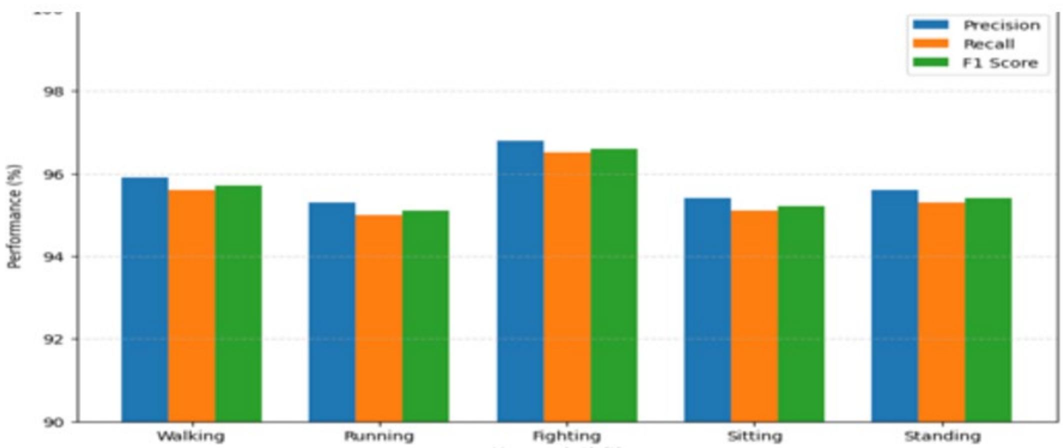
6.3. ACTIVITY RECOGNITION RESULTS

Table 3

Table 3 Evaluation Metrics for Human Activity Recognition

Activity	Precision (%)	Recall (%)	F1 Score (%)
Walking	95.9	95.6	95.7
Running	95.3	95.0	95.1
Fighting	96.8	96.5	96.6
Sitting	95.4	95.1	95.2
Standing	95.6	95.3	95.4
Overall	**95.8**	**95.5**	**95.6**

YOLOv8 combined with Deep SORT is used to classify and track human activities such as Walking, Running, Fighting, Sitting, and Standing.



The system effectively differentiates normal and abnormal behavior with minimal misclassification.

6.4. ACTIVITY–EMOTION FUSION

A rule-based fusion strategy combines activity and emotion information to reduce false alarms. For example:

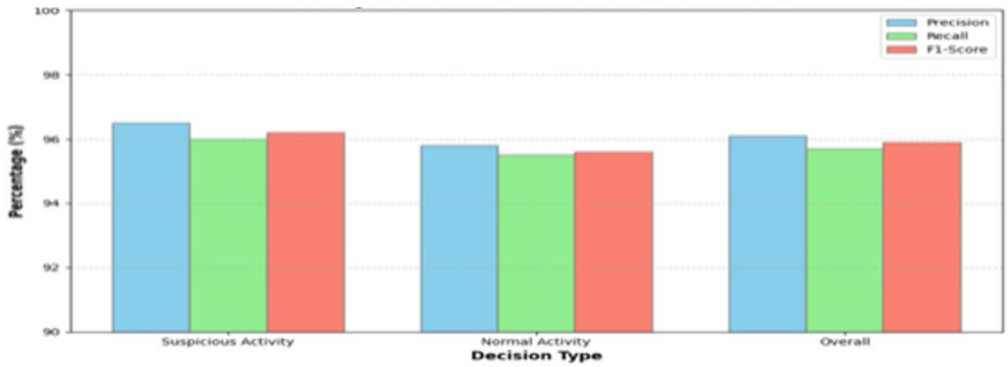
- Running + Neutral/Happy → Normal
- Running + Fear/Anger → Suspicious
- Fighting + Any emotion → Suspicious

Table 4

Table 4 Evaluation of Fused Activity–Emotion Classification

Decision Type	Precision (%)	Recall (%)	F1 Score (%)
Suspicious Activity	96.5	96.0	96.2
Normal Activity	95.8	95.5	95.6
Overall	96.1	95.7	95.9

The fusion approach improves overall system reliability, achieving an overall accuracy of ~96%.



6.5. SYSTEM-LEVEL OUTPUT

- Real-time bounding box detection for humans and faces
- Simultaneous display of activity labels and emotion states
- Alert generation for abnormal behavior scenarios
- Event logging for future analysis

6.6. EDGE SYSTEM PERFORMANCE

- Entire system runs on Raspberry Pi without cloud dependency
- Low latency inference suitable for real-time surveillance
- Reduced bandwidth usage due to edge processing
- Enhanced privacy and deployability in public environments

6.7. RESULTS OBTAINED

The system consists of a real-time web-based GUI developed using the Flask framework and HTML, CSS, and JavaScript technologies for visualizing results in terms of video streaming and AI results.

The frontend (HTML/CSS/JS) displays real-time video feed with bounding boxes and labels (activities and emotions) that are recognized in real-time.

The backend (Flask) processes video frames using YOLOv8 (object detection), DeepSORT (object tracking), CNN (emotion recognition), and BiLSTM (activity recognition).

Real-time outputs are transferred from backend to frontend using REST APIs/socket communication techniques.

System Output Attributes:

- Live video stream from web dashboard
- Bounding boxes of detected persons (using YOLOv8 + DeepSORT)
- Labels on detected emotions (using CNN models)
- Labels of activities such as Normal/Suspicious
- Real-time fusion-based decision making display
- Pop-up notification when an abnormal activity occurs
- Option to turn on buzzer output using GPIO of RPI

Decision Making Criteria (Fusion Based):

- Running person with neutral/happy face → Normal activity
- Running person with fear/anger → Suspicious activity
- Fighting with any emotion → High Risk Alert

Edge Computing:

All operations occur at edge computing, i.e., Raspberry Pi through Flask server by guaranteeing:

- Minimum latency real-time performance
- Local and secure computation
- Zero-cloud computing dependency

Output Visualization

Figure 2 GUI Home Screen

Figure 3 Real-time Emotion Detection Output

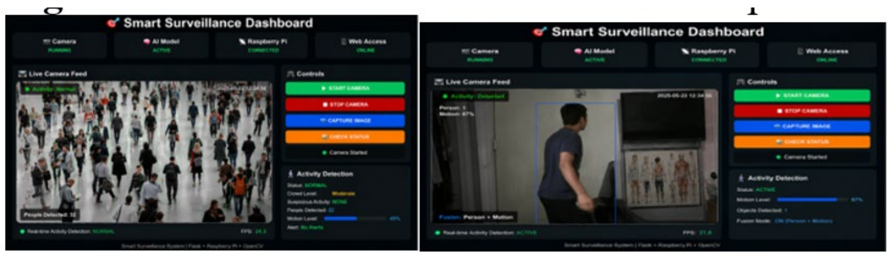


Figure 4 Normal Activity Detection

7. CONCLUSION

The developed system successfully utilizes the object detector YOLOv8, multi-object tracker DeepSORT, facial emotion recognizer CNN, and behavior analyzer BiLSTM. Moreover, the system successfully utilizes decision-level fusion to integrate activity data and emotion data for improving accuracy in detecting any form of abnormal behavior. Real-time processing with low latency is achieved by running the application on a Raspberry Pi. Experiments prove that the system has high accuracy with regard to precision, recall, and F1-Score. In addition, the system can operate in a real-time manner. Moreover, the system has a convenient user interface based on the Flask framework.

The future work will involve the improvement of the efficiency of the models by applying the models of lightweight transformers, improving detection accuracy in extreme lighting and occlusion conditions, and applying the system to the environment of multi-camera distributed surveillance with cloud-edge computing hybrid approach.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

A. Kuzdeuov, M. Zakaryanov, A. Tleuliyev, and H. A. Varol, “OpenThermalPose2: Extending the Open-Source Annotated Thermal Human Pose Dataset With More Data, Subjects, and Poses,” IEEE Transactions on Biometrics, Behavior, and Identity Science, vol. 7, no. 4, pp. 902–913, Oct. 2025, doi: 10.1109/TBIOM.2025.3575499.

A. M. Mekala et al., “Fused YOLO and Traditional Features for Emotion Recognition From Facial Images of Tamil and Russian Speaking Children: A Cross-Cultural Study,” IEEE Access, vol. 13, pp. 86828–86840, 2025, doi: 10.1109/ACCESS.2025.3569771.

A. Rehman, T. Saba, M. Z. Khan, R. Damaševičius, and S. A. Bahaj, “Internet-of-Things-Based Suspicious Activity Recognition Using Multimodalities of Computer Vision for Smart City Security,” Security and Communication Networks, vol. 2022, Art. no. 8383461, 2022.

B. Ma et al., “Distracted Driving Behavior and Driver’s Emotion Detection Based on Improved YOLOv8 With Attention Mechanism,” IEEE Access, vol. 12, pp. 37983–37994, 2024, doi: 10.1109/ACCESS.2024.3374726.

Eissa, M. E. A. (2025). The Influence of AI-Generated Content on Trust and Credibility Within Specialized Online Communities: A Brief Review on Proposed Conceptual Framework. ShodhAI: Journal of Artificial Intelligence, 2(2), 1–14. https://doi.org/10.29121/shodhai.v2.i2.2025.40

I. U. Khan, S. Afzal, and J. W. Lee, “Human Activity Recognition via Hybrid Deep Learning Based Model,” Sensors, vol. 22, no. 1, p. 323, 2022.

K. U. Duja, I. A. Khan, and M. Alsuhaibani, “Video Surveillance Anomaly Detection: A Review on Deep Learning Benchmarks,” IEEE Access, vol. 12, pp. 164811–164842, 2024, doi: 10.1109/ACCESS.2024.3491868.

M. M. A. Parambil et al., “Navigating the YOLO Landscape: A Comparative Study of Object Detection Models for Emotion Recognition,” IEEE Access, vol. 12, pp. 109427–109442, 2024, doi: 10.1109/ACCESS.2024.3439346.

M. M. Zaidi et al., “Suspicious Human Activity Recognition From Surveillance Videos Using Deep Learning,” IEEE Access, vol. 12, pp. 105497–105510, 2024, doi: 10.1109/ACCESS.2024.3436653.

M. Q. Gandapur and E. Verdú, “ConvGRU-CNN: Spatiotemporal Deep Learning for Real-World Anomaly Detection in Video Surveillance System,” International Journal of Interactive Multimedia and Artificial Intelligence, vol. 8, no. 4, p. 88, 2023.

N. Dwivedi, D. Singh, and D. Kushwaha, “A Novel Approach for Suspicious Activity Detection With Deep Learning,” Multimedia Tools and Applications, vol. 82, pp. 1–24, 2023, doi: 10.1007/s11042-023-14445-7.

O. Saidani et al., “An Efficient Human Activity Recognition Using Hybrid Features and Transformer Model,” IEEE Access, vol. 11, pp. 101373–101386, 2023, doi: 10.1109/ACCESS.2023.3314492.

P. Kuppusamy and V. C. Bharathi, “Human Abnormal Behaviour Detection Using CNNs in Crowded and Uncrowded Surveillance – A Survey,” Measurement: Sensors, vol. 24, p. 100510, 2022.

Q. U. A. Arshad et al., “Anomalous Situations Recognition in Surveillance Images Using Deep Learning,” Computers, Materials & Continua, vol. 76, no. 1, pp. 1103–1125, 2023.

R. R. Subramanian and V. Vasudevan, “A Deep Genetic Algorithm for Human Activity Recognition Leveraging Fog Computing Frameworks,” Journal of Visual Communication and Image Representation, vol. 77, p. 103132, 2021, doi: 10.1016/j.jvcir.2021.103132.

R. Rajeswari, “Anomalous Human Activity Recognition From Video Sequences Using BRISK Features and Convolutional Neural Networks,” Galaxy International Interdisciplinary Research Journal, vol. 10, no. 2, pp. 269–280, 2022.

R. Vrskova, R. Hudec, P. Kamencay, and P. Sykora, “A New Approach for Abnormal Human Activities Recognition Based on ConvLSTM Architecture,” Sensors, vol. 22, no. 8, p. 2946, 2022.

S. Khanam, M. Sharif, X. Cheng, and S. Kadry, “Suspicious Action Recognition in Surveillance Based on Handcrafted and Deep Learning Methods: A Survey of the State of the Art,” Computers and Electrical Engineering, vol. 120, p. 109811, 2024.

T. Saba et al., “Suspicious Activity Recognition Using Proposed Deep L4-Branched-ActionNet With Entropy Coded Ant Colony System Optimization,” IEEE Access, vol. 9, pp. 89181–89197, 2021, doi: 10.1109/ACCESS.2021.3091081.

W. Ullah et al., “CNN Features With Bi-Directional LSTM for Real-Time Anomaly Detection in Surveillance Networks,” Multimedia Tools and Applications, vol. 80, no. 11, pp. 16979–16995, 2021.

Y. Qin et al., “ORAF-YOLO: A Lightweight Factory Unsafe Behavior Recognition Method With Occlusion-Robust Adaptive Fusion,” IEEE Access, 2026, doi: 10.1109/ACCESS.2026.3687749.