iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IX Month of publication: September 2025

DOI: https://doi.org/10.22214/ijraset.2025.74044

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IX Sep 2025- Available at www.ijraset.com

# ThreatGaze: Smart Visual Behavior Analytics for Risk Prevention

Dr. Girish Kumar D[1], Miss. Thela Shirisha[2]
[1]Professor & HOD, Department of MCA, Ballari Institute of Technology and Management, Ballari, Karnataka, India
[2]Department of MCA, Ballari Institute of Technology and Management, Ballari, Karnataka, India

***Abstract: This paper presents a Visual Behavior Analysis (VBA) framework designed to detect and interpret human activity. integrates YOLOv8 for human detection, MediaPipe Pose for posture recognition, identification. Together, these tools enable accurate differentiation between normal and suspicious behaviors. Experimental evaluation shows behavior recognition accuracy of 85–95% while maintaining 25–30 frames per second, making it suitable for real-time monitoring. The proposed approach enhances response time, supports proactive crime prevention, and provides a scalable platform for safer public environments.***

## I. INTRODUCTION

Visual Behavior Analysis (VBA) is a growing field within artificial intelligence and computer vision that aims to automatically interpret human gestures, movements, and expressions. Unlike traditional monitoring, which relies on manual observation and is prone to fatigue and inaccuracy, AI-driven surveillance enables faster and more precise analysis of live video streams. With urbanization and population growth, the shortcomings of conventional systems have become more visible, particularly in managing large amounts of real-time video data. Current solutions struggle with challenges such as poor lighting, occlusion, and limited datasets, which reduce their reliability. This work proposes a framework that integrates YOLOv8 for human detection, MediaPipe Pose for pose estimation, and Haar Cascade for facial recognition, delivering an efficient and real-time solution for security monitoring.

## II. RELATED WORKS

Human activity recognition has been widely studied in recent years, with approaches evolving from handcrafted features to deep learning methods. Traditional systems often failed under real-world conditions such as lighting changes, occlusions, or crowded backgrounds. These methods deliver strong accuracy but often demand.

Another direction explored is gaze-based analysis, where eye movement patterns such as fixation and pupil dilation are used to estimate stress levels and attention. While this approach adds value in specialized applications, results vary greatly depending on the user and environment, reducing consistency. Graph-based models, such as GCNs, have been applied to extract relationships between key body points, achieving encouraging results but with challenges in maintaining real-time performance.

In the context of crowded public spaces, several studies have explored abnormal behavior detection using spatiotemporal deep learning techniques. Although these systems can highlight unusual movements, they frequently raise false alarms due to background noise and variations in human actions.

### *A. Methodological Insights*

1) Action Recognition: Using CNNs, LSTMs, or Transformers to classify video-based human activities.
2) Gaze Tracking: Estimating attention and stress from eye movement data.
3) Emotion Analysis: Detecting affective states by combining body and facial cues through models like GCNs.
4) Integrated Detection: Leveraging YOLO-based object detection with pose estimation for identifying both normal and suspicious behaviors in real time.

### *B. Findings from Literature*

Research consistently reports recognition accuracy ranging from 85% to 95% under controlled setups. Object detection systems like YOLO have proven effective in identifying threat-related actions such as vandalism, while gaze-based models provide useful behavioral signals despite limited robustness. Emotion recognition works well in structured environments but often slows down when applied to live scenarios.wed down under real-time requirements.

### *C. Common Limitations*

Despite these advances, the following limitations remain:

*1)* Performance drops in low light, occluded views, or dense crowds.
*2)* False positives caused by natural variations in human movements.
*3)* Limited training datasets, restricting generalization to real-world environments.

## III. PROPOSED METHOD

*1)* Video Input & Preprocessing: Captures and enhances frames from surveillance feeds for improved clarity.
*2)* Human Detection: YOLOv8 is employed to identify multiple individuals within a scene.
*3)* Pose Estimation: MediaPipe Pose maps skeletal landmarks such as arms, legs, and torso to interpret body movements.
*4)* Alerts & Notifications: Detected threats generate real-time alerts through visual overlays or message notifications.

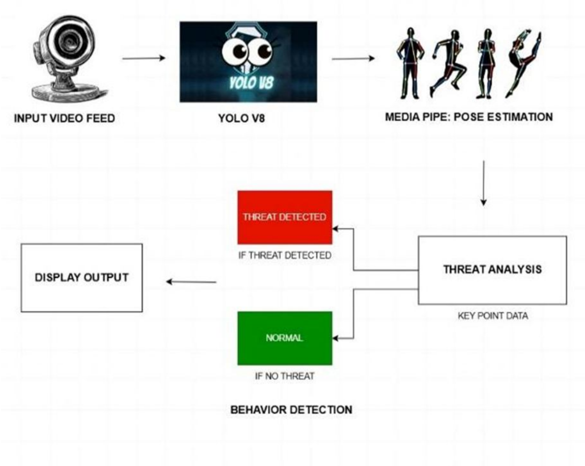


Figure 1 Schematic diagram of the proposed model

Additionally, the framework is built with scalability in mind, making it adaptable for future enhancements, such as integration with advanced healthcare systems, face recognition, and multilingual support

### *A. Architecture*

To ensure accurate classification of human behavior using computer vision and AI. The main components are:

*1)* Real-Time Video Processing: A live video feed is captured through a camera and processed using YOLOv8, which detects human subjects with high accuracy. This ensures efficient tracking of individuals in dynamic environments.
*2)* Pose Estimation & Threat Analysis**:** Once human subjects are detected, MediaPipe Pose Estimation extracts skeletal key points to analyse body movements. These movements are then assessed by a Threat Analysis Module**,** which classifies behaviors based on predefined threat indicators such as aggressive postures or sudden movements.
*3)* Behavior Classification & Output Display**:** Based on the threat analysis, the system categorizes behavior as "Threat Detected" (red alert) or "Normal" (green status). The results are displayed in real-time, enabling security personnel or automated systems to respond immediately.

This architecture enables real-time behavioral assessment for applications in security, workplace monitoring, and human activity recognition.

### *B. Workflow*

Our real-time behavioral analysis system is designed to efficiently detect and classify human behavior from live video feeds, ensuring quick and accurate assessments. The system follows a structured workflow to identify potential threats and enhance security in various environments.:

*1)* Frame Preprocessing: Each frame is preprocessed to improve image quality and prepare it for analysis.
*2)* Human Detection: YOLOv8 detects humans in the frame and generates bounding boxes around them.
*3)* Pose Estimation: MediaPipe Pose estimates key body landmarks and classifies the pose.

4) Behavior Analysis: The system analyses the extracted features to classify behavior as normal or threatening.
5) Real-Time Alerts: If a threat is detected, the system generates alerts and overlays visual indicators on the video feed.
6) Data Logging: All detected behaviors are logged in a centralized database for subsequent analysis and reporting.

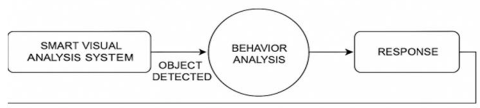


Figure 2 Level 0 Data flow diagram

This structured approach makes the system efficient, scalable, and highly reliable for security monitoring, workplace safety, and behavioral analysis. By combining state-of-the-art computer vision methods combined with real-time processing, it ensures a proactive and effective security solution.

## IV. RESULTS AND DISCUSSION

### A. *Comparative Analysis*

System testing demonstrated improved performance compared to conventional CCTV monitoring. The integration of YOLOv8 and Media Pipe Pose provided accurate classification of diverse behaviors in different environments. Unlike manual systems, automated recognition reduced response time and improved security efficiency

### B. *Evaluation Metrics*

1) Accuracy: Classification accuracy ranged from 85% to 95%, depending on environmental conditions.
2) Threat Detection**:** Aggressive or abnormal actions were flagged with approximately 90% confidence.

### C. *Comparison with State-of-the-Art Systems*

Unlike traditional motion-based detection, which struggles in complex scenarios, the inclusion of pose estimation improved the system's ability to distinguish between harmless gestures and actual threats.

System Requirements

1) *Hardware Requirements*

a) Processor : A multi-core CPU (e.g., Intel i5 or equivalent) to handle real-time video processing tasks.
b) GPU: NVIDIA GTX 1060 or higher for efficient deep learning computations and model inference.
c) RAM: Minimum 8 GB (16 GB recommended for smoother performance).
d) Storage: SSD with at least 500 GB capacity for storing video feeds, system logs, and datasets.
e) Camera: High-resolution IP cameras (1080p or higher) to capture clear live video feeds.

*Network Requirements:*

A high-speed internet connection to support real-time data transmission and remote monitoring.
Minimum bandwidth of 10 Mbps for uninterrupted video streaming and data transfer.

2) *Software Requirements*

a) Operating System: Ubuntu 20.04 LTS or Windows 10/11 (64-bit) for system compatibility and stability.
b) Frameworks and Libraries:

YOLOv8: Utilized for real-time human detection in video streams.
MediaPipe Pose: Employed for accurate pose estimation and body landmark detection.
OpenCV: Used for video processing, frame extraction, and image enhancement.
Haar Cascade Classifiers: Integrated for facial recognition tasks.

*c)* Additional Tools

- TensorFlow/PyTorch: For training and deploying deep learning models.
- CUDA and cuDNN: Essential for GPU acceleration during deep learning operations.
- Database: MySQL or PostgreSQL for storing detected behaviors, logs, and system data.

*d)* Dependencies

Python libraries such as NumPy, Pandas, Matplotlib, and Flask for data manipulation, visualization, and API integration.

## V. CONCLUSION

This research introduces a Visual Behavior Analysis system that integrates YOLOv8, MediaPipe Pose, and Haar Cascade for real-time human behavior monitoring. The framework achieves high accuracy, reliable response speed, and practical applications in public safety, healthcare, and workplace monitoring. While the system addresses many limitations of manual surveillance, challenges remain in scalability, false positives, and handling subtle behaviors. Future improvements will include sensor fusion, larger datasets, and multilingual support for wider applicability.

## REFERENCES

[1] Vattikunta Mahitha, Allenki Usha Reddy, Jangili Sunitha, Dr. P. Rama "Detection of Human Behavior and Abnormality Using YOLO and Conv2D," International Journal of Scientific Development and Research (IJSDR), Vol. 8, Issue 4, April 2023, pp. 1009-1016.

[2] Hieu H. Pham, Louahdi Khoudour, Alain Crouzil, Pablo Zegers, Sergio A. Velastin* "A Review on Deep Learning Approaches for Video-Based Human Action Recognition," arXiv preprint arXiv:2208.03775, 2022.

[3] Riku Arakawa, Kiyosu Maeda, Hiromu Yakura "Providence: A Machine Learning-Based Multimodal Tool for Analyzing Conversational Behavior," ACM, 2024.

[4] Joseph Redmon, Ali Farhadi "YOLOv3: Enhancing Real-Time Object Detection," arXiv preprint arXiv:1804.02767, 2018.arXiv preprint arXiv:1804.02767, 2018.

[5] I. Goodfellow, J. Shlens, C. Szegedy* "Harnessing and Explaining Adversarial Examples in Deep Learning," arXiv preprint arXiv:1412.6572, 2015.

[6] Paul Viola, Michael J. Jones "Real-Time Face Detection for Surveillance Applications,"

[7] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp "Human Pose Recognition from a Single Depth Image,"

[8] Yan LeCun, Yoshua Bengio, Geoffrey Hinton "A Comprehensive Review of Deep Learning Techniques," Nature, 2015.

[9] Sepp Hochreiter, Jürgen Schmidhuber* "Long Short-Term Memory Networks for Sequence Learning," Neural Computation, 1997.

[10] Marco Cristani, R. Raghavendra, Alessio Del Bue, Vittorio Murino "Understanding Human Behavior in Surveillance Through Social Signal Processing," Neurocomputing, 2011.