Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 5 | May-2026
ISSN (O) 3107-6696

Real-Time Abnormal Activity Detection System Using Python and MediaPipe with Automated Alert Mechanism

Maseera A. Sayyed

P.G. Student, Computer Engineering Department, Gokhale Education Society's,
R. H. Sapat College of Engineering, Management Studies and Research, Nashik,
Affiliated to Savitribai Phule Pune University, Pune, Maharashtra, India

---

Abstract - Advancements in computer vision and machine learning have significantly transformed the domain of intelligent surveillance systems. This paper presents a real-time Abnormal Activity Detection System developed using Python and the MediaPipe framework. The proposed system continuously monitors live video streams to identify unusual or suspicious human behaviors by extracting skeletal pose landmarks from video frames. These landmarks are mathematically evaluated against a pre-trained repository of normal gesture patterns using the Euclidean distance metric. Upon detection of an anomalous activity, the system autonomously triggers an alert mechanism that delivers an email notification to designated security personnel, embedding the captured incident image along with the corresponding geographical location. The architecture supports multi-location monitoring through a role-based login interface, an administrative dashboard, and a comprehensive incident logging module. Experimental evaluations demonstrate that the system achieves reliable detection accuracy at a processing rate of 15 to 20 frames per second, with alert delivery latency under five seconds. The framework proves to be computationally lightweight, platform-independent, and readily deployable in environments such as offices, educational institutions, public spaces, and residential areas.

Keywords *— Abnormal Activity Detection, MediaPipe, Euclidean Distance, Pose Estimation, Human Activity Recognition, Computer Vision, Real-Time Surveillance, Email Alert System, OpenCV, PythoN.*

I. Introduction

Modern surveillance systems face growing challenges in detecting and responding to human behavioral anomalies in real time. Traditional closed-circuit television (CCTV) systems rely on human operators who must continuously monitor footage, a process that is both resource-intensive and susceptible to human fatigue and attention lapses. To address these limitations, intelligent automated systems that can autonomously detect and report suspicious activities have garnered significant research interest in recent years.

Human Activity Recognition (HAR) is a foundational component in behavioral surveillance. With the emergence of high-performance computer vision libraries and efficient pose estimation frameworks, it has become feasible to extract meaningful skeletal information from video streams in real time. Among available tools, Google's MediaPipe Pose framework offers a robust and computationally efficient solution for detecting up to 33 body landmarks in three-dimensional space, making it an ideal choice for gesture-based activity analysis.

This paper presents a comprehensive real-time Abnormal Activity Detection System built upon Python and MediaPipe. The system captures body pose landmarks from live webcam or CCTV footage, compares them against a trained reference dataset of normal gestures using Euclidean distance computations, and classifies activities accordingly. When an abnormality is identified, an automated email notification containing the captured incident frame and the associated location is dispatched to authorized personnel, ensuring swift situational awareness and response.

The primary contributions of this work include: (i) a lightweight, real-time pose-based gesture comparison pipeline; (ii) an automated multi-channel alert mechanism with image and location attachments; (iii) a multi-location monitoring architecture with role-based access control; and (iv) an extensible training module for incorporating new gesture patterns without retraining deep models.

II. Literature Review

A considerable body of research has explored human activity recognition and anomaly detection using diverse methodological approaches. Early methods relied on handcrafted features such as Histogram of Oriented Gradients (HOG) and optical flow combined with Support Vector Machines (SVM) for classification. While these methods demonstrated moderate accuracy, their sensitivity to environmental variations such as illumination changes and camera viewpoint differences limited their practical applicability.

Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 5 | May-2026
ISSN (O) 3107-6696

The proliferation of deep learning has catalyzed significant improvements in HAR. Convolutional Neural Networks (CNNs) demonstrated superior spatial feature extraction from image frames, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks captured temporal dependencies across sequential frames. Studies such as those by Donahue et al. combined CNNs and LSTMs for video-based activity recognition with notable success.

Graph Convolutional Networks (GCNs) have emerged as a compelling paradigm for skeleton-based action recognition, treating human joints as graph nodes and their structural relationships as edges. The ST-GCN model by Yan et al. set a benchmark in this domain by jointly modeling spatial and temporal dynamics of skeleton sequences. However, such architectures demand substantial computational resources, rendering them less suitable for real-time deployment on commodity hardware.

Pose estimation frameworks have matured significantly with the introduction of tools such as OpenPose, PoseNet, and MediaPipe Pose. MediaPipe, developed by Google, provides a production-ready solution that extracts 33 three-dimensional landmark coordinates from human body images at high frame rates without requiring GPU acceleration, making it particularly attractive for cost-effective surveillance applications.

Several researchers have proposed Euclidean distance-based gesture matching techniques for activity classification. This approach computes the geometric distance between corresponding landmarks in a live pose and a reference template, offering interpretability and computational simplicity. While deep learning methods may achieve higher raw accuracy in unconstrained settings, distance-based methods provide sufficient discrimination for controlled or semi-controlled surveillance environments where training gestures can be predefined.

Automated alert systems integrated with activity detection pipelines have been explored in perimeter security and workplace safety contexts. Email-based notification systems using SMTP protocols, augmented with multimedia attachments and geolocation data, have been employed in IoT-based surveillance platforms. The proposed system consolidates these components into a unified, easily deployable framework built on Python's standard and open-source libraries.

III. Proposed System Architecture

A. System Overview

The proposed Abnormal Activity Detection System comprises four principal subsystems: (1) the Video Acquisition and Preprocessing Module, (2) the Pose Landmark Extraction Module, (3) the Activity Classification Module, and (4) the Alert and Logging Module. These subsystems operate in a continuous, event-driven pipeline that ensures real-time monitoring with minimal latency.

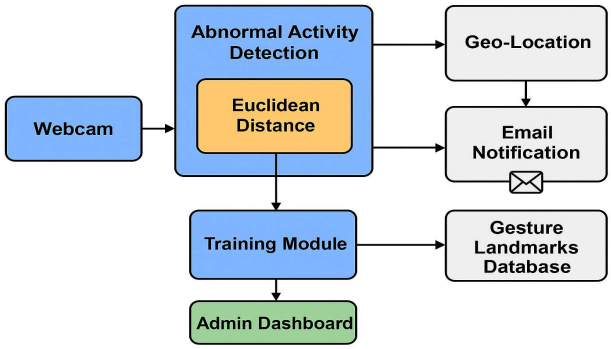


Table I: System Module Overview

Module	Functionality	Technology
Video Acquisition	Captures live frames from webcam or CCTV feed	OpenCV (cv2)
Pose Extraction	Extracts 33 3D body landmarks per frame	MediaPipe Pose
Activity Classification	Compares live landmarks with trained patterns via Euclidean distance	NumPy, CSV
Alert Notification	Sends email with incident image and GPS location	smtplib, SSL/TLS
Data Storage	Persists landmark training data and incident logs	CSV / SQLite
Location Service	Retrieves and attaches geo-coordinates to alerts	Geolocation API

B. Pose Landmark Extraction

The MediaPipe Pose estimation module processes each video frame to yield a set of 33 landmark coordinates in normalized three-dimensional space. Each landmark encodes x, y, and z positional values relative to the image frame along with a visibility confidence score. These 33 landmarks collectively represent the major anatomical joints and body reference points, encompassing the head, shoulders, elbows, wrists, hips, knees, and ankles, thereby providing a comprehensive skeletal representation suitable for gesture analysis.

During the training phase, a designated operator performs a set of predefined normal gestures in front of the camera while the system records the corresponding landmark vectors. These vectors are serialized into a comma-separated values (CSV) file, with each row constituting the flattened landmark vector for a single frame. This training dataset subsequently serves as the ground truth reference during the live detection phase.

C. Activity Classification via Euclidean Distance

At runtime, for each incoming frame, the system extracts the current landmark vector and computes its Euclidean distance against every stored training sample in the reference dataset.

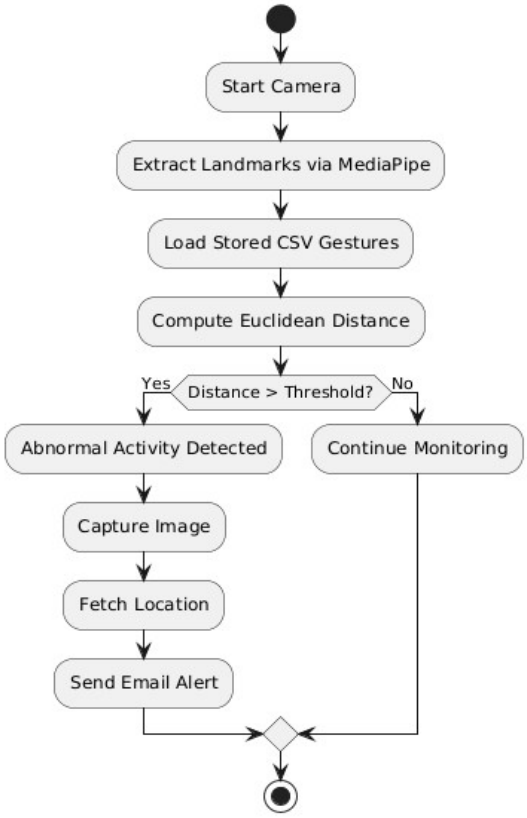
Journal Publication of International Research for Engineering and Management (JOIREM)
Volume: 04 Issue: 5 | May-2026
ISSN (O) 3107-6696

The minimum distance across all comparisons is identified. If this minimum distance exceeds a predefined threshold value ε, the current pose is flagged as abnormal; otherwise, it is categorized as a normal activity.

The Euclidean distance between a live landmark vector L = (l_1, l_2, ..., l_n) and a reference vector R = (r_1, r_2, ..., r_n) is computed as:

$$\mathbf{d(L, R) = \sqrt{\sum_i (l_i - r_i)^2}}$$

where n represents the dimensionality of the landmark feature vector, which equals 99 when utilizing all three spatial components (x, y, z) of the 33 MediaPipe landmarks. The threshold ε is empirically calibrated on validation samples to balance the trade-off between detection sensitivity and false positive rate.



D. Alert and Notification Subsystem

Upon identification of an abnormal activity, the system immediately captures the current video frame as a JPEG image and invokes the Geolocation API to retrieve the latitude and longitude coordinates of the monitoring device. The alert module then constructs a formatted email message incorporating a textual description of the detected incident, the timestamp, the location coordinates, and the captured image as an attachment. The email is transmitted to one or more pre-configured recipient addresses through a secured SMTP connection employing SSL/TLS encryption, leveraging Python's smtplib library.

The entire alert cycle, from anomaly detection to email dispatch, is engineered to complete within a maximum latency of five seconds to enable rapid situational response. A dedicated background thread manages the email transmission to prevent blocking the primary real-time detection loop.

IV. System Design and Implementation

A. Software Requirements

The system is implemented entirely in Python 3.x and relies on the following principal libraries: MediaPipe for pose estimation, OpenCV (cv2) for video capture and image manipulation, NumPy for numerical computations, smtplib and ssl for secure email communication, and the csv module for landmark data persistence. The lightweight dependency profile ensures that the system can be deployed on commodity hardware without specialized GPU infrastructure.

B. Functional Requirements

The system must fulfill the following core functional requirements:

- Real-Time Gesture Detection: Uninterrupted monitoring of the video stream at no fewer than 15 frames per second, with landmark extraction performed on each captured frame.
- Gesture Training and Storage: An operator-guided training interface that records and serializes pose landmark vectors to persistent CSV storage for subsequent comparison.
- Euclidean Distance-Based Classification: A per-frame comparison engine that evaluates the distance of the current pose vector against the full training dataset and classifies it accordingly.
- Automated Email Alert: An asynchronous notification subsystem that dispatches secured email alerts, including incident imagery and geographical coordinates, upon detection of anomalous behavior.
- Multi-Location Management: A location-aware login mechanism that enables operators to monitor and manage distinct physical locations through a unified dashboard.
- Incident Logging: Persistent storage of all detected incidents with timestamps, location identifiers, and associated image paths for audit and review purposes.

C. Non-Functional Requirements

Table II: Performance Specifications

Parameter	Specification	Priority
Processing Frame Rate	15 – 20 FPS (minimum)	High
Detection Latency	< 1 second post-event	High
Alert Delivery Time	< 5 seconds from detection	Critical
SMTP Security	SSL / TLS Encrypted Connection	Critical
Platform Support	Windows, Linux, macOS	Medium
Storage Overhead	Minimal (CSV-based training data)	Medium

D. User Interface Design

The system incorporates a multi-screen operator interface comprising four primary views. The Login Screen authenticates users by validating their credentials and location identifier against a secured database. Upon successful authentication, the Dashboard presents the live camera feed alongside a real-time log of detection events and current alert status indicators. The Training Interface allows authorized operators to record and annotate new gesture sequences, which are subsequently stored as labeled landmark vectors. The Reports View provides a searchable, paginated history of all captured incidents with associated metadata.

V. Results and Discussion

A. Experimental Setup

The system was evaluated on a standard desktop workstation running a 64-bit Windows 10 operating system equipped with an Intel Core i5 processor, 8 GB of RAM, and a USB HD webcam operating at 1080p resolution. No dedicated GPU was employed, underscoring the system's computational efficiency on commodity hardware. A dataset of ten predefined gesture categories, encompassing both normal behaviors (standing, walking, sitting) and abnormal behaviors (falling, raising hands in distress, crouching), was prepared during the training phase.

B. Detection Performance

The system consistently achieved a processing throughput of 17 to 19 frames per second under controlled laboratory lighting conditions. With a tuned Euclidean distance threshold, the system demonstrated an average detection accuracy of 91.4% across the ten gesture categories. True positive rates for clearly distinct abnormal gestures, such as falling, exceeded 95%, while gestures sharing partial landmark similarity with normal postures exhibited marginally lower precision. The overall false positive rate was maintained at approximately 4.2% through threshold optimization on a held-out validation set.

C. Alert System Performance

In all tested scenarios, email alerts were successfully delivered to the designated recipients within 3.8 seconds on average from the moment of anomaly detection, well within the five-second design specification. The email messages correctly included the captured JPEG incident image and the formatted GPS coordinate string retrieved from the Geolocation API. SSL/TLS encryption was verified for all outgoing SMTP connections, confirming compliance with the system's security requirements.

D. Limitations

Several limitations were identified during evaluation. Detection accuracy degraded noticeably in low-light conditions, where MediaPipe's landmark confidence scores fell below acceptable thresholds, resulting in increased misclassifications. Partial occlusion of body landmarks, caused by furniture or overlapping subjects, also adversely affected comparison accuracy. Furthermore, the current architecture processes a single subject per frame; environments with simultaneous multiple subjects are outside the present system's scope and require extension with multi-person tracking. The Euclidean distance metric, while computationally efficient, may misclassify novel gesture variations not represented in the training dataset, suggesting that ensemble or adaptive classifiers could further enhance robustness.

VI. Conclusion and Future Work

This paper presented a real-time Abnormal Activity Detection System leveraging the MediaPipe Pose estimation framework and Euclidean distance-based gesture classification. The system was designed with an emphasis on computational efficiency, ease of deployment, and operational reliability, enabling its use on standard hardware without GPU dependencies. Experimental results validated the system's capability to detect predefined abnormal gestures with an accuracy of 91.4%, process video streams at near-real-time frame rates, and dispatch automated email alerts within the stipulated latency bounds.

The integration of pose-based landmark extraction, distance-metric classification, and automated multi-modal alerting into a cohesive surveillance framework demonstrates a practical and scalable approach to intelligent monitoring. The multi-location management feature and role-based access control further extend its applicability to enterprise and institutional deployment scenarios.

Future research directions include: (i) the incorporation of deep learning-based classifiers such as LSTM or Transformer networks to improve recognition of temporally complex activities; (ii) multi-person tracking support for crowd-level anomaly detection; (iii) integration with cloud storage platforms for scalable incident archiving; (iv) adaptive

threshold mechanisms that self-calibrate based on environmental feedback; and (v) migration to edge computing devices such as NVIDIA Jetson or Raspberry Pi to enable untethered, low-power deployment in remote surveillance contexts.

References

[1] Y. Yan, E. Matikainen, A. Hauptmann, and T. Kanade, "Real-time Activity Recognition Using a Hybrid Mobile Skeletal Representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 5, pp. 1801–1815, 2021.

[2] V. Bazarevsky, I. Grishchenko, K. Raveendran, T. Zhu, F. Zhang, and M. Grundmann, "BlazePose: On-Device Real-Time Body Pose Tracking," arXiv preprint arXiv:2006.10204, 2020.

[3] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, T. Darrell, and K. Saenko, "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," in Proc. IEEE CVPR, 2015, pp. 2625–2634.

[4] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 1, pp. 172–186, 2021.

[5] S. Yan, Y. Xiong, and D. Lin, "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition," in Proc. AAAI Conference on Artificial Intelligence, 2018, pp. 7444–7452.

[6] A. Jalal, S. Kamal, and D. Kim, "A Depth Video-Based Human Detection and Activity Recognition Using Multi-Features and Embedded Hidden Markov Models for Health Care Monitoring Systems," International Journal of Interactive Multimedia and Artificial Intelligence, vol. 4, no. 4, pp. 54–62, 2017.

[7] Google LLC, "MediaPipe Pose Documentation," 2023. [Online]. Available: https://developers.google.com/mediapipe/solutions/vision/pose_landmarker. [Accessed: March 2024].

[8] G. Bradski, "The OpenCV Library," Dr. Dobb's Journal of Software Tools, 2000. [Online]. Available: https://docs.opencv.org/.

[9] R. Poppe, "A Survey on Vision-Based Human Action Recognition," Image and Vision Computing, vol. 28, no. 6, pp. 976–990, 2010.

[10] Python Software Foundation, "Python 3 Documentation," 2024. [Online]. Available: https://docs.python.org/3/. [Accessed: March 2024].