IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

SentinelPose: A Temporal Consistency-Aware Framework for Pose-Based Behavioral Anomaly Detection in Retail Surveillance

Madhuri Borawake[1], Ashwini Chandel[2], Aishwarya Godase[3], Amol Chaudhari[4], Vikas Vighne[5]

Professor Department of Computer Engineering, PDEA's College of Engineering Manjari Bk. Hadapsar, Pune, Maharashtra[1]

Students Department of Computer Engineering, PDEA's College of Engineering Manjari Bk. Hadapsar, Pune, Maharashtra[2-5]

Abstract: Retail stores widely use CCTV cameras to monitor customer activities and maintain security. However, continuously watching multiple video feeds is difficult for security staff, and unusual activities may sometimes go unnoticed. Existing automated surveillance methods can help identify suspicious movements, but their performance may be affected by inaccurate body pose estimation, tracking errors, and normal customer actions that appear unusual. To address these challenges, this paper proposes SentinelPose, a reliability-aware framework for detecting unusual human behavior in retail environments. The proposed system uses YOLO to detect people in CCTV footage and ByteTrack to track their movements across video frames. MediaPipe Pose is used to extract body landmarks, which are organized into time-based sequences and analyzed using an LSTM Autoencoder. The model learns common movement patterns and identifies possible anomalies based on reconstruction error. To improve the reliability of alerts, the proposed approach considers the quality of extracted pose information and checks whether unusual movements continue over multiple frames before generating an alert. When a potential anomaly is identified, the system is designed to save relevant video evidence and display the alert on a monitoring dashboard for security staff to review. The framework is intended to reduce unnecessary alerts and support efficient surveillance without using facial recognition or automatically identifying anyone as a shoplifter. The proposed method will be evaluated using detection precision, recall, false alarm rate, and processing speed.

Keywords: Pose-Based Behavioral Anomaly Detection, Retail Surveillance, Human Movement Analysis, Unsupervised Sequence Learning, LSTM Autoencoder, Multi-Person Tracking, Real-Time Video Analytics, Suspicious Activity Alerting.

1. INTRODUCTION

1.1. Background:

Closed-Circuit Television (CCTV) cameras are widely used in retail stores to monitor customer activities, protect merchandise, and maintain security. These cameras are usually placed at store entrances, billing counters, product shelves, and storage areas. However, most surveillance systems still depend on security staff to watch video footage and identify unusual activities. Monitoring multiple cameras for long periods is difficult, particularly in busy stores where many customers move around at the same time.

Recent advances in computer vision and deep learning have made it possible to develop smarter surveillance systems. These technologies can detect people, estimate body poses, and analyze movements over time. Human pose estimation focuses on identifying important body points, such as the shoulders, elbows, hips, knees, and ankles. By studying how these points change across video frames, a system can learn movement patterns and identify activities that differ from normal behavior.

Pose-based analysis offers a useful approach to studying human movement in retail environments. However, unusual movements do not always indicate suspicious behavior. Therefore, an effective surveillance system should consider both the movement itself and the reliability of the information extracted from the video.

1.2. Problem Statement

Identifying unusual customer behavior from CCTV footage is challenging because many normal shopping activities may appear suspicious when viewed without sufficient context. For example, bending down to pick up an item, reaching for

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

a product on a shelf, adjusting clothing, or placing an item in a shopping basket are common activities. An automated system may incorrectly classify these movements as anomalies, leading to unnecessary security alerts.

The quality of human pose estimation creates another challenge. In crowded stores, a person's body may be partially hidden by shelves, products, or other customers. As a result, some body landmarks may be missing or incorrectly detected. Changes in camera angles and interruptions in person tracking can further affect the accuracy of movement analysis.

These problems may cause an anomaly detection system to produce unreliable results. In particular, a short-lived movement or an error in pose estimation may be mistaken for an unusual activity. Therefore, a more reliable approach is needed to evaluate the quality of extracted pose information and examine movement patterns over time before generating an alert.

This research proposes SentinelPose, a reliability-aware temporal framework designed to identify potential behavioral anomalies in retail surveillance videos. The framework aims to reduce unnecessary alerts by considering pose quality and the persistence of anomaly scores. Its purpose is to assist security personnel in reviewing unusual activities, not to automatically determine whether a person has committed shoplifting.

1.3 Motivation:

The main motivation for this research is to improve the reliability of automated behavior analysis in retail surveillance. Although modern computer vision techniques can detect people and analyze their movements, their performance may be affected by incomplete pose information, tracking errors, and variations in normal customer behavior.

A single unexpected movement is not enough to conclude that an activity is suspicious. For instance, a customer may bend suddenly to pick up a dropped item or stretch to reach a product on a high shelf. If the system generates an alert for every unusual movement, security staff may receive too many false alarms. This can increase their workload and make important events more difficult to identify.

To address this problem, SentinelPose focuses on two main factors: the reliability of the detected body landmarks and the consistency of unusual movement patterns across consecutive video frames. The framework is intended to consider both factors when deciding whether an event should be flagged for review.

The study aims to investigate whether this combined approach can reduce unnecessary alerts while maintaining the ability to identify meaningful behavioral anomalies. By linking detected events with relevant video evidence, the proposed system also aims to make the review process more practical for security personnel.

1.4. Research Objectives:

The main objective of this research is to propose and evaluate a reliability-aware temporal framework for detecting potential behavioral anomalies in retail surveillance videos.

The specific objectives are as follows:

1. To detect people in retail CCTV footage using YOLO and track their movements across video frames using ByteTrack.
2. To extract human body landmarks using MediaPipe Pose and normalize the landmark coordinates for movement analysis.
3. To represent human movements as sequences of pose information and analyze these sequences using an LSTM Autoencoder trained on normal movement patterns.
4. To assess pose reliability by examining landmark visibility and the availability of valid body landmark observations.
5. To develop a temporal verification mechanism that checks whether elevated anomaly scores continue across multiple observations before generating an alert.
6. To develop an alert and evidence management module that stores relevant information for human review.
7. To evaluate the proposed framework using suitable performance measures, including precision, recall, false alarm frequency, processing latency, and computational requirements.

1.5 Main Research Contributions:

This paper introduces SentinelPose, a proposed framework that combines human pose estimation, multi-object tracking, and temporal anomaly detection for retail surveillance. The main contributions of the proposed research are described below.

1. Integrated Pose-Based Behavioral Analysis: The framework combines YOLO-based person detection, ByteTrack tracking, MediaPipe Pose estimation, and an LSTM Autoencoder to analyze human movement patterns across video frames.
2. Pose Reliability Assessment: A reliability assessment mechanism is proposed to evaluate the quality of extracted body landmarks. It considers landmark visibility and missing observations to identify pose sequences that may be unsuitable for reliable movement analysis.
3. Temporal Consistency Verification: The framework introduces a temporal verification mechanism that examines anomaly scores across multiple observations. This approach is intended to reduce alerts caused by brief movement changes and temporary pose estimation errors.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

4. Reliability-Aware Alert Generation: The proposed alert mechanism considers both pose reliability and the persistence of anomaly scores before reporting a potential event. Relevant video segments are intended to be stored or linked to the alert so that security personnel can review the event.
5. Practical Performance Evaluation: The study defines an experimental approach for evaluating detection performance, false alarm frequency, processing latency, and computational requirements. The evaluation is intended to examine the framework's suitability for retail surveillance systems, including environments with limited computing resources.

These contributions describe the proposed design and evaluation plan. Their effectiveness must be demonstrated through experiments and comparisons with suitable baseline methods before any performance improvement can be claimed.

2.LITERATURE REVIEW AND RELATED WORK

2.1 YOLO-Based Shoplifting Detection Systems :

Automated shoplifting detection has become an important research area because manually monitoring CCTV footage requires continuous attention from security staff. Recent surveillance systems use object detection and machine learning techniques to identify people, observe their movements, and recognize activities that may require further investigation.

YOLO (You Only Look Once) is widely used in computer vision because it can detect objects in video frames with relatively low processing time. In retail surveillance, YOLO-based models can identify customers and provide bounding boxes that help locate individuals in a camera view. These detections can then be used for tracking and behavior analysis.

Perumal and Maha Ganesh [1] presented a shoplifting detection approach using YOLOv5, MediaPipe Pose, and machine learning-based behavior analysis. Their work shows how object detection and body-pose information can be combined to study customer activities. However, identifying people and recognizing movements alone may not be sufficient to determine whether an activity is genuinely suspicious.

The proposed SentinelPose framework extends this general approach by examining movement patterns over time using an LSTM Autoencoder. It also proposes pose reliability assessment and temporal consistency verification to investigate whether alerts can be made more dependable.

2.2 Human Pose-Based Anomaly Detection :

Human pose estimation is a computer vision technique that identifies important body points, including the shoulders, elbows, wrists, hips, knees, and ankles. These points represent a person's body posture and can be used to understand movements across video frames.

Pose-based anomaly detection focuses on changes in body movement rather than relying only on visual appearance. This is useful in surveillance because clothing, lighting, and background conditions may vary between people and locations.

Hirschorn and Avidan [2] introduced a human pose anomaly detection method using normalizing flows. Their approach studies patterns in human pose sequences to identify movements that differ from normal activities. The research demonstrates that skeletal movement information can support video anomaly detection without depending entirely on image appearance.

Rashvand et al. [3] also investigated pose-based anomaly detection in retail environments through the PoseLift dataset and benchmark. Their work provides a useful foundation for studying unusual customer behavior using body-pose information.

However, pose estimation can become less reliable when people are partially hidden, overlap with others, or move outside the camera view. Such situations may produce incomplete body landmarks and affect anomaly detection.

SentinelPose proposes a pose reliability assessment mechanism to examine the quality of extracted landmarks before using them for behavioral analysis.

2.3 Deep Learning for Temporal Behavior Analysis :

Human activities cannot always be understood from a single video frame. Many actions involve a sequence of movements that develop over time. For example, reaching toward a shelf, bending to pick up an item, and placing a product into a shopping basket may look different in individual frames but form a normal shopping activity when observed together.

Deep learning methods can analyze these movement sequences and identify patterns that may not be visible through frame-by-frame observation.

Karim et al. [4] proposed a real-time weakly supervised video anomaly detection method that focuses on identifying unusual events without waiting for an entire video to be processed. Their research highlights the importance of timely anomaly detection in surveillance applications.

Zhang et al. [5] investigated multi-scale video anomaly detection using spatio-temporal representation learning. Their method examines information across different spatial and temporal levels to understand activity patterns.

These studies show that analyzing movement over time is important for identifying unusual behavior. However, a system must also consider how quickly an anomaly can be detected and whether temporary changes should generate alerts.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

SentinelPose uses fixed-length pose sequences to study movement patterns and proposes temporal verification to check whether an anomaly continues across multiple observations.

2.4 LSTM Autoencoder-Based Anomaly Detection :

Long Short-Term Memory (LSTM) networks are designed to process sequential information and learn relationships between observations over time. They are useful for analyzing human movements because body positions change continuously across video frames.

An LSTM Autoencoder contains two main parts: an encoder and a decoder. The encoder learns a compact representation of an input sequence, while the decoder attempts to reconstruct the original sequence. The difference between the input and reconstructed output is called reconstruction error.

When an Autoencoder is trained mainly on normal movement patterns, unfamiliar movements may produce higher reconstruction errors. These errors can be used as anomaly scores, although a high score does not always indicate genuinely suspicious behavior.

Şengönül et al. [6] explored surveillance anomaly detection using LSTM autoencoding and local-minima assistance. Their study demonstrates the use of temporal reconstruction for identifying unusual events in surveillance videos.

This approach is relevant to SentinelPose because the proposed framework also uses an LSTM Autoencoder to learn normal movement patterns. However, SentinelPose uses extracted human pose sequences rather than depending entirely on full video-frame information.

The proposed framework further investigates whether pose reliability assessment and temporal consistency verification can help reduce unnecessary alerts caused by incomplete observations or short-term movement variations.

2.5 Multi-Person Detection and Tracking :

Retail environments often contain several customers moving at the same time. To analyze each person's behavior separately, a surveillance system must detect individuals and follow their movements across consecutive video frames.

Object detection identifies the location of a person in a frame, while multi-object tracking connects detections belonging to the same individual over time. Tracking systems normally assign temporary IDs to organize the movement history of each detected person.

ByteTrack is a multi-object tracking method that associates detected objects across frames, including detections with lower confidence scores. This can help maintain tracks when detection confidence changes.

Wang and Mariano [7] investigated a tracking framework combining YOLOv8s and ByteTrack. Their work supports the use of object detection and tracking techniques for following multiple individuals in video sequences.

However, tracking becomes difficult when customers overlap, move behind shelves, or leave and re-enter the camera view. Incorrect tracking may mix the movement information of different individuals and affect later anomaly analysis.

SentinelPose proposes using YOLO and ByteTrack to maintain separate pose sequences for tracked customers. The framework also considers tracking interruptions when deciding whether enough reliable movement information is available for anomaly detection.

2.6 Existing Retail Surveillance Datasets :

Datasets play an important role in training and evaluating behavioral anomaly detection systems. A suitable dataset should contain examples of normal activities and, where possible, annotated unusual events that can be used for testing.

Common video anomaly detection datasets include CUHK Avenue, ShanghaiTech, and UCF-Crime. These datasets are useful for studying general surveillance anomalies, but their environments and activities do not always represent everyday retail shopping behavior.

Rashvand et al. [3] introduced PoseLift, a retail-focused dataset and benchmark designed to support research on pose-based shoplifting detection. It provides a more directly relevant starting point for investigating human movement anomalies in retail environments.

For SentinelPose, dataset selection must consider the availability of pose sequences, original video footage, annotations, recording conditions, and permission for research use. Pose-only data may be suitable for evaluating the anomaly detection model but may not support testing the complete video-to-alert pipeline.

The proposed study will therefore examine dataset compatibility before training and will separate training, validation, and testing data by recording or session to reduce the risk of data leakage.

2.7 Comparative Analysis of Existing Research :

Previous research has contributed important methods for person detection, human pose estimation, temporal behavior analysis, and surveillance anomaly detection. However, these studies address different parts of the overall problem.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Research study	Main contribution	Relevance to SentinelPose	Limitation / research opportunity
[1]YOLOv5-Based Shoplifting Detection	Combines object detection, pose estimation, and machine learning	Supports retail activity analysis	Opportunity to study temporal anomaly verification
[2]Normalizing Flows for Human Pose Anomaly Detection	Uses pose sequences and normalizing flows	Supports pose-based anomaly detection	Does not specifically focus on retail alert management
[3]PoseLift	Introduces a real-world retail pose dataset and benchmark	Directly supports retail anomaly research	Provides a benchmark rather than the complete proposed SentinelPose application
[4]Real-Time Weakly Supervised Video Anomaly Detection	Studies timely anomaly detection	Supports low-latency evaluation	Uses a different supervision and feature-learning approach
[5]Multi-Scale Video Anomaly Detection	Analyzes video patterns at different spatial and temporal scales	Supports temporal behavior analysis	Not specifically designed around SentinelPose's pose reliability mechanism
[6]LSTM Autoencoding for Surveillance	Uses temporal reconstruction for anomaly detection	Supports LSTM-based anomaly scoring	Does not specifically investigate the proposed pose-quality gating mechanism
[7]YOLOv8s and ByteTrack Tracking	Combines object detection with multi-object tracking	Supports customer movement tracking	Does not directly address retail behavioral anomaly decisions

YOLO-based approaches support person detection, while tracking algorithms help maintain movement histories across video frames. Pose-based methods provide body movement representations, and temporal deep learning models analyze changes in these representations over time.

Retail-focused research such as PoseLift provides useful datasets and evaluation methods. Meanwhile, LSTM-based approaches demonstrate how reconstruction error can support anomaly scoring.

The proposed SentinelPose framework combines these established ideas and investigates an additional reliability-aware temporal verification mechanism. The aim is to examine how pose quality and anomaly persistence influence alert generation in retail surveillance.

Table I summarizes the main contributions, relevance, and limitations of the reviewed studies.

2.8 Research Gap :

Existing studies have shown that computer vision and deep learning can support the detection of unusual human activities in surveillance videos. However, practical retail environments create challenges that may affect the reliability of automated anomaly alerts.

One challenge is the quality of extracted human pose information. Customers may be partially hidden by shelves, shopping carts, or other people, resulting in missing or inaccurate body landmarks. If these observations are directly used for anomaly detection, they may produce misleading anomaly scores.

Another challenge is the interpretation of short-term unusual movements. Normal activities, such as bending, stretching, reaching for products, or adjusting bags, may temporarily differ from the movement patterns learned by a model. Generating alerts from individual high anomaly scores can increase unnecessary notifications.

A further challenge is maintaining useful detection performance while controlling processing time and alert delays, especially on computers with limited hardware resources.

These observations motivate the need to investigate how pose reliability and temporal consistency can be considered together during anomaly detection. Although related studies have explored pose-based analysis and temporal modeling, the specific combination proposed in this research requires direct comparison with existing approaches before its effectiveness or novelty can be established.

To investigate these issues, SentinelPose proposes a reliability-aware temporal framework that combines pose-quality assessment, LSTM Autoencoder reconstruction error, and temporal anomaly verification. The framework is designed to generate potential-anomaly alerts with supporting video evidence for human review.

The main research objective is to evaluate whether the proposed mechanisms can reduce unnecessary alerts while maintaining useful detection sensitivity and acceptable processing latency in retail surveillance environments.

3. PROPOSED SYSTEM ARCHITECTURE

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

3.1 Overview of the Proposed SentinelPose Framework:

SentinelPose is a proposed intelligent surveillance framework designed to identify unusual human movements in retail environments. The main purpose of the system is to support security personnel by automatically analyzing CCTV footage and highlighting activities that may require further attention. Instead of directly identifying a customer as a shoplifter, the framework focuses on detecting movement patterns that differ from normal shopping behavior.

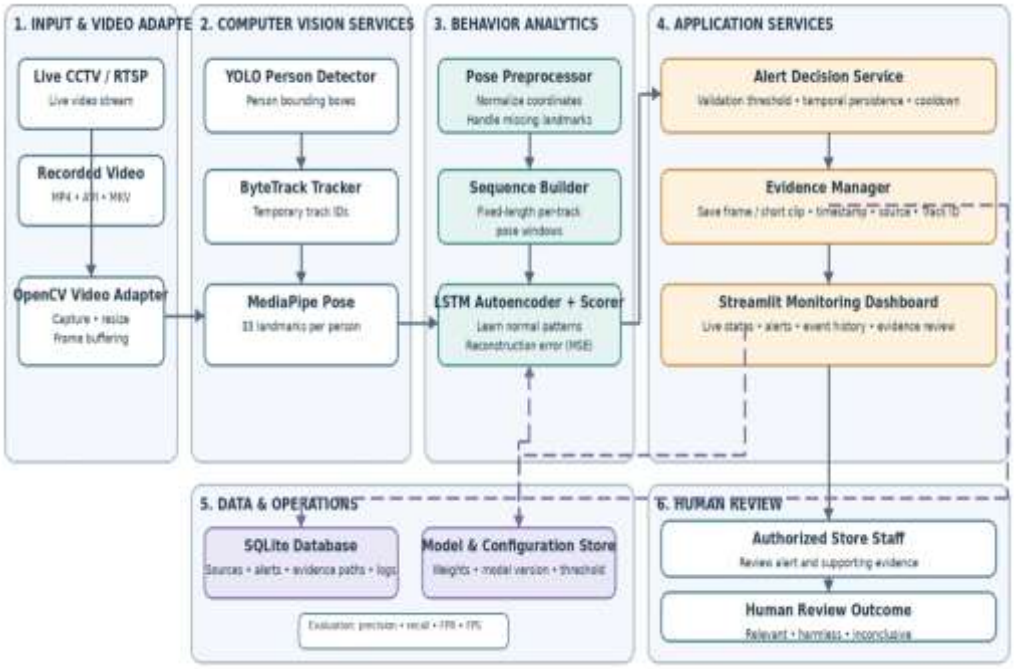
The proposed system combines computer vision and deep learning techniques to analyze human activities over time. It uses YOLO for person detection, ByteTrack for tracking individuals, MediaPipe Pose for extracting body landmarks, and an LSTM Autoencoder for identifying unusual movement patterns. A reliability-aware temporal verification mechanism is also proposed to examine the quality of pose information and the consistency of anomaly scores before generating alerts.

The framework follows a step-by-step processing approach. Video frames are collected from CCTV footage, people are detected and tracked, and their body movements are represented using pose landmarks. These landmarks are arranged into time-based sequences and analyzed to identify possible behavioral anomalies. When the defined alert conditions are satisfied, the system is designed to generate a notification and save relevant video evidence for human review.

The proposed architecture is modular, allowing each component to be developed and evaluated separately. This design also supports future improvements without requiring major changes to the complete system.

3.2 System Architecture Diagram :

The proposed architecture of SentinelPose is shown in Fig. 1. The system consists of several connected modules that work together to analyze CCTV footage and identify unusual human movements. It begins with video acquisition, followed by person detection using YOLO and tracking using ByteTrack. MediaPipe Pose extracts body landmarks, which are normalized and arranged into temporal sequences. An LSTM Autoencoder analyzes these sequences to calculate anomaly scores. The proposed reliability assessment and temporal verification mechanisms help determine whether an alert should be generated. Finally, the system stores supporting evidence and displays potential anomalies through a monitoring dashboard for human review.



3.2 Proposed system architecture of SentinelPose.

3.3 Video Acquisition and Frame Preprocessing :

The video acquisition module is responsible for collecting footage from supported surveillance sources. The initial framework is designed to process recorded CCTV videos in common formats such as MP4, AVI, and MKV. Live CCTV streams using protocols such as RTSP may also be supported, depending on the available hardware and implementation requirements.

OpenCV is proposed for reading video frames and performing basic preprocessing operations. Each frame is extracted from the input video and prepared for the next processing stage. Preprocessing may include resizing frames, adjusting

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

image dimensions, and converting color formats according to the requirements of the detection and pose estimation models.

Frame resolution and processing frequency can affect both detection quality and computational requirements. Therefore, the framework is intended to use consistent preprocessing settings during training, validation, and testing.

The video acquisition module also records timestamps and monitors the availability of the input source. These details help connect detected events with the correct video footage during alert generation and evidence storage.

3.4 YOLO-Based Person Detection :

The person detection module uses the YOLO object detection algorithm to identify people present in each video frame. YOLO processes an image and predicts bounding boxes around detected objects along with confidence scores.

In SentinelPose, only detections belonging to the person class are selected for further processing. Each detected person is represented by a bounding box containing the approximate location of the individual in the video frame.

The bounding boxes are passed to the tracking module, which uses them to follow people across consecutive frames. A lightweight YOLO model is proposed to reduce computational requirements, particularly when the system operates on computers without dedicated graphics hardware.

The main responsibility of this module is to locate people accurately. It does not analyze whether a person's activity is normal or unusual. Behavioral analysis is performed in later stages using the extracted body movements.

3.5 ByteTrack Multi-Object Tracking :

After detecting people in each frame, SentinelPose uses ByteTrack to follow their movements over time. ByteTrack is a multi-object tracking algorithm that connects detections across consecutive video frames and assigns temporary tracking IDs.

These tracking IDs allow the system to maintain a separate movement history for each detected person. For example, when multiple customers are present in a store, the tracking module helps organize their pose information individually rather than combining their movements.

ByteTrack uses detection confidence and association techniques to maintain object tracks. Its ability to consider lower-confidence detections can help preserve tracking continuity when detection confidence temporarily decreases.

However, tracking may become difficult when people overlap, move behind shelves, or leave the camera view. Such interruptions can affect the quality of movement sequences.

The proposed framework therefore considers tracking continuity when preparing pose sequences. When tracking information becomes unreliable, the affected observations may be excluded or handled according to predefined rules to avoid producing misleading anomaly scores.

3.6 MediaPipe Human Pose Estimation :

The pose estimation module uses MediaPipe Pose to identify important body landmarks for each tracked person. Human pose estimation represents the position of different body parts, allowing the system to study posture and movement without depending entirely on a person's visual appearance.

MediaPipe Pose provides 33 body landmarks, including points associated with the head, shoulders, elbows, wrists, hips, knees, and ankles. These landmarks describe the approximate structure of the human body and help capture changes in posture across video frames.

For each tracked individual, the proposed framework applies pose estimation to the corresponding person region. The extracted landmarks are then associated with the temporary tracking ID assigned by ByteTrack.

This process allows SentinelPose to build individual movement histories and analyze how body positions change over time.

Pose estimation may become less accurate when body parts are hidden or when a person is positioned at an unusual angle. Therefore, the proposed framework also considers landmark visibility information when determining whether the extracted pose is reliable enough for further analysis.

3.7Pose Feature Extraction and Normalization :

After estimating the human pose, the framework converts body landmarks into numerical features suitable for machine learning.

Each MediaPipe Pose landmark is represented using four values:

- x: Horizontal position of the landmark.
- y: Vertical position of the landmark.
- z: Relative depth information.
- v: Landmark visibility value.

Since MediaPipe Pose provides 33 landmarks, the total number of features for one frame is calculated as:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

$F = 33 \times 4 = 132$

Therefore, each valid pose observation is represented by a feature vector containing 132 values.
Before using these features for anomaly detection, normalization is applied to reduce the influence of differences in person position and body size within the video frame. A consistent normalization method helps the model focus on movement patterns rather than absolute image coordinates.

The proposed framework also includes pose reliability assessment. Landmark visibility values are examined to estimate the quality of extracted pose information. Observations containing too many missing or low-confidence landmarks may be marked as unreliable.
Depending on the selected preprocessing strategy, incomplete observations may be handled through interpolation, masking, or rejection. The same preprocessing rules must be followed during model training and evaluation.
This reliability assessment is intended to reduce the influence of poor-quality pose information on later anomaly detection stages.

3.8 Temporal Sequence Construction :
Human behavior develops over time, so analyzing individual frames may not provide enough information to understand a movement. To address this limitation, SentinelPose organizes pose features into temporal sequences.
The proposed framework initially uses a sequence length of 30 consecutive frames. Each frame contains 132 pose features, resulting in a sequence represented as:
$X \in \mathbb{R}^{(30 \times 132)}$
Here, 30 represents the number of frames, while 132 represents the pose features extracted from each frame.
A sliding-window approach is proposed to create movement sequences for each tracked person. As new frames become available, the sequence window moves forward and includes the latest pose observations.
Each sequence is associated with its corresponding temporary tracking ID to prevent movement information from different people from being mixed.
The framework also considers pose reliability and tracking continuity during sequence construction. If a sequence contains too many unreliable observations or significant tracking interruptions, it may be excluded from anomaly analysis.
The initial sequence length of 30 frames is a design choice and will be examined during experimentation. Different sequence lengths may capture different movement patterns and affect detection performance and processing time.

3.9 LSTM Autoencoder-Based Anomaly Detection
The anomaly detection module uses an LSTM Autoencoder to learn common human movement patterns and identify possible behavioral deviations.
An LSTM Autoencoder consists of an encoder and a decoder. The encoder processes an input pose sequence and learns a compact representation of its temporal information. The decoder then attempts to reconstruct the original sequence using the learned representation.
The proposed model is intended to be trained mainly on normal shopping movements, such as walking, browsing products, reaching toward shelves, and placing items into shopping baskets.
During testing, incoming pose sequences are passed through the trained Autoencoder. The difference between the original sequence and the reconstructed sequence is measured using Mean Squared Error (MSE).
The reconstruction error is expressed as:

$$E(X) = (1 / TF) \Sigma \Sigma (X - \hat{X})^2$$

Where:
- X: Original pose sequence.
- $\hat{X}$: Reconstructed pose sequence.
- T: Number of frames in the sequence.
- F: Number of pose features per frame.

A lower reconstruction error suggests that the observed movement is similar to patterns learned during training. A higher error indicates that the movement differs from those patterns.
However, a high reconstruction error does not automatically mean that a person is performing a suspicious or harmful activity. Normal but uncommon movements may also produce high anomaly scores.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

To address this limitation, SentinelPose proposes a reliability-aware temporal verification mechanism. The mechanism considers the quality of pose observations and examines whether elevated anomaly scores persist across multiple valid sequences before generating an alert.
The anomaly threshold and temporal verification settings will be selected using validation data and evaluated through controlled experiments.

3.10 Alert Generation and Evidence Management :
The alert generation module is responsible for identifying events that satisfy the defined anomaly detection and temporal verification conditions.
Instead of generating an alert immediately after a single high anomaly score, the proposed framework checks whether the detected deviation continues across multiple reliable observations. This process is intended to reduce unnecessary alerts caused by brief movements or temporary errors in pose estimation.
When the required conditions are satisfied, the system generates a notification labelled "Potential Anomaly – Review Required."
Each alert is intended to contain important information, including:

- Alert ID and timestamp.
- Video source or camera information.
- Temporary tracking ID.
- Anomaly score and selected threshold.
- Model version and alert status.
- Location of the supporting image or video clip.

The evidence management module saves relevant frames or video segments associated with the detected event. These records are linked to the corresponding alert using a database identifier.
SQLite is proposed for storing alert details, evidence references, and review information. A Streamlit-based dashboard provides an interface through which authorized security personnel can view alerts, examine supporting footage, and record review decisions.
The dashboard is intended to distinguish automatically generated anomaly alerts from observations confirmed by human reviewers.
SentinelPose does not use facial recognition to identify customers and does not automatically declare that shoplifting has occurred. Its purpose is to highlight unusual movement patterns that may require further examination.
By combining pose-based anomaly detection, reliability assessment, temporal verification, and evidence management, the proposed framework aims to support more practical and responsible retail surveillance. Its actual effectiveness will need to be established through implementation and experimental evaluation.

3.11 System Activity Diagram :
The activity diagram shown in Fig. 2 represents the step-by-step workflow of the proposed SentinelPose system. The process begins when a video source is selected and the system starts reading video frames. Each frame is processed to detect people, track their movements, and extract body-pose information. The extracted features are prepared for anomaly detection, and the LSTM Autoencoder calculates an anomaly score. The system then checks the reliability of pose observations and verifies whether unusual movement patterns continue over multiple observations. If the defined conditions are satisfied, an alert is generated and supporting evidence is stored. Otherwise, the system continues processing the next frames. This workflow illustrates how the system processes video information and makes alert decisions.

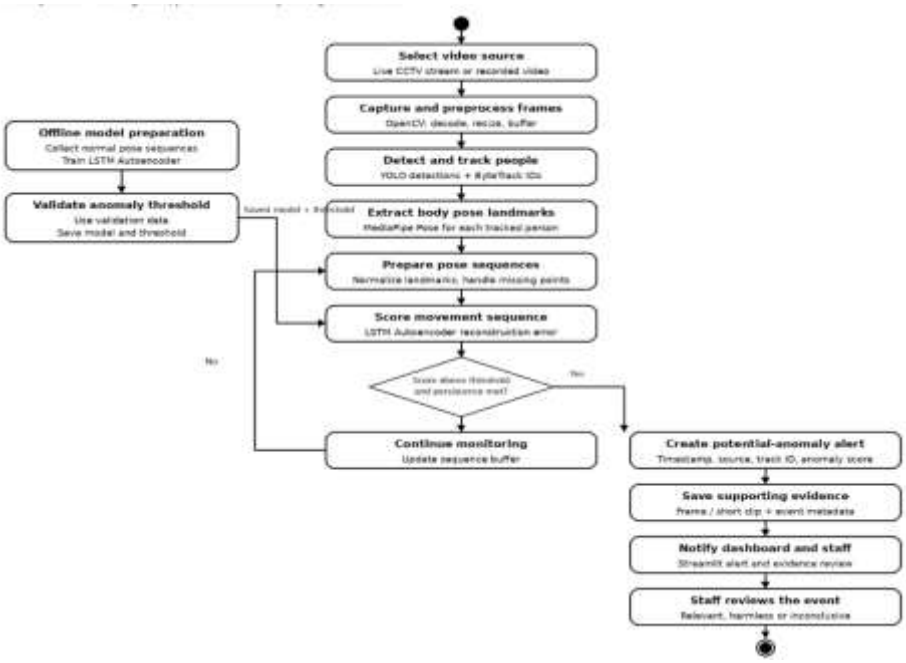
© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065



3.11Activity Diagram

3.12 System Use Case Diagram :

The use case diagram presented in Fig. 3 illustrates the main interactions between users and the proposed SentinelPose system. The primary actors are the security personnel and system administrator. Security personnel can monitor video processing, view generated alerts, examine supporting evidence, and update the review status of detected events. The administrator is responsible for managing system settings, video sources, and authorized access. Automated operations, such as person detection, tracking, pose extraction, anomaly analysis, and alert generation, are performed internally by the system. The diagram provides an overview of the proposed system's main functions and user responsibilities.

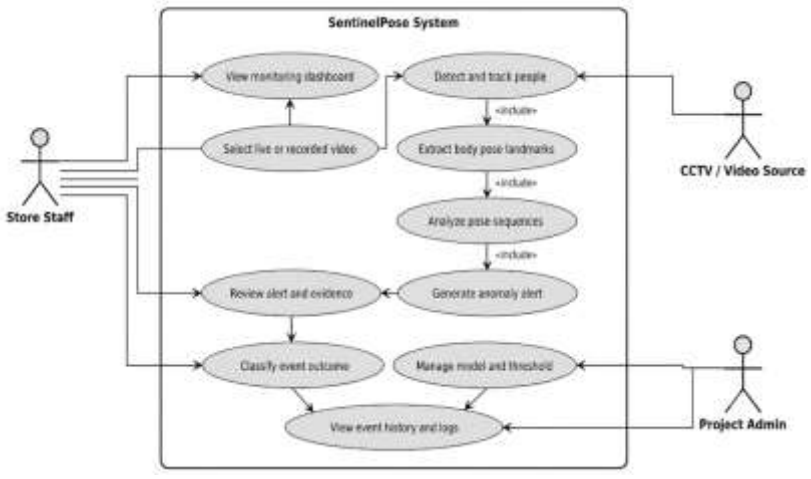


Fig 3.12 Use Case Diagram

4. PROPOSED RELIABILITY-AWARE TEMPORAL METHODOLOGY

4.1 Overview of the Proposed Method:

The proposed SentinelPose framework is designed to identify unusual human movements in retail surveillance videos by analyzing body-pose information over time. The main idea is to examine not only how a person's movement differs from normal behavior but also how reliable the extracted pose information is and whether the unusual movement continues for a meaningful period.

Traditional anomaly detection approaches may generate high anomaly scores when a person makes an uncommon movement or when the input data contains errors. In retail environments, these situations can occur because of blocked

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

camera views, incomplete body landmarks, tracking interruptions, or normal customer activities such as bending and reaching for products.

To investigate these challenges, SentinelPose proposes a reliability-aware temporal methodology. The system first detects and tracks people in video frames and extracts their body landmarks. These landmarks are normalized and arranged into fixed-length sequences. An LSTM Autoencoder then analyzes the sequences and calculates reconstruction errors to identify possible deviations from learned movement patterns.

The proposed method also includes a pose reliability assessment mechanism that examines landmark visibility and the availability of valid observations. A temporal consistency mechanism checks whether elevated anomaly scores continue across multiple reliable sequences before an alert is generated.

The overall objective is to investigate whether combining pose reliability assessment with temporal anomaly verification can reduce unnecessary alerts while maintaining useful detection performance.

4.2 Human Pose Feature Representation :

Human pose estimation helps identify the positions of different body parts and understand how a person moves over time. In the proposed SentinelPose framework, MediaPipe Pose is used to extract 33 body landmarks for each tracked person in a video frame. These landmarks represent important body points, including the shoulders, elbows, wrists, hips, knees, and ankles.

Each landmark is described by four values: horizontal coordinate (x), vertical coordinate (y), relative depth (z), and visibility score (v). The pose information of a person at frame $t$can be represented as:

$$P_t = \{(x_{t,i}, y_{t,i}, z_{t,i}, v_{t,i})\}_{i=1}^{33}$$

Here, $t$represents the current video frame, and $i$represents the landmark index.

Since each of the 33 landmarks contains four values, the total number of features extracted from one frame is:

$$F = 33 \times 4 = 132$$

Thus, each frame produces a pose feature vector containing 132 values, provided that the pose estimation output is available.

Before further analysis, the extracted pose features are normalized to reduce the effects of differences in a person's position and apparent size in the camera view. This allows the model to focus more on body movement patterns rather than relying only on the person's absolute position in the image.

After normalization, the pose feature vectors are arranged into sequences of consecutive video frames. These sequences are then used by the LSTM Autoencoder to learn common movement patterns and identify deviations from the learned patterns. To ensure consistent results, the same feature extraction and normalization procedures are applied during training, validation, and testing.

One technical correction: Your original text states that $F = 33 \times 4 = 13$, but the correct result is **132**. I have corrected it in the rewritten version.

Also, for stronger technical accuracy in your conference paper, specify the exact normalization method you use, such as torso-centered coordinates or scaling relative to shoulder or hip distance. Only describe a method that you actually implement.

4.3 Pose Reliability Assessment:

The quality of human pose estimation can vary depending on lighting conditions, camera angles, body movements, and visibility. For example, a customer's arm may be hidden behind a shelf, or the lower part of the body may be outside the camera's view. In such cases, some body landmarks may be missing or estimated with low confidence.

When the extracted pose information is incomplete or inaccurate, it may not correctly represent the person's actual movement. If these unreliable observations are directly used by the anomaly detection model, they may lead to incorrect reconstruction errors and unnecessary alerts.

To reduce this problem, the proposed SentinelPose framework includes a pose reliability assessment mechanism. This mechanism uses the landmark visibility values provided by MediaPipe Pose to estimate the quality of the detected pose in each video frame.

The average visibility score for frame $t$is calculated as follows:

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

$$Q_t = \frac{1}{N}\sum_{i=1}^{N} v_{t,i}$$

where:

- $Q_t$represents the average visibility score of the pose in frame t.
- $N = 33$represents the total number of body landmarks.
- $v_{t,i}$represents the visibility value of landmark $i$in frame t.

A higher value of $Q_t$indicates that the landmarks generally have higher visibility scores, whereas a lower value suggests that some body parts may be partially hidden or difficult to estimate.
Based on this score, the pose observation is classified using a predefined reliability threshold:

$$R_t = \begin{cases} 1, & Q_t \geq \gamma \\ 0, & Q_t < \gamma \end{cases}$$

Here, γrepresents the reliability threshold selected using validation data. When $R_t = 1$, the pose satisfies the selected visibility requirement. When $R_t = 0$, the pose is marked as potentially unreliable and requires appropriate handling before further analysis.

This mechanism provides an initial way to identify pose observations that may contain insufficient visual information. However, average visibility alone cannot guarantee that all landmarks are correctly positioned or that the extracted pose accurately represents the person's movement. Therefore, the effectiveness of this reliability measure must be evaluated experimentally to determine how well it supports anomaly detection and reduces unnecessary alerts.
Technical suggestion for your conference paper: Explain what SentinelPose does when $R_t = 0$. For example, does it skip the frame, reduce its contribution to the anomaly score, or wait for more reliable observations? Clearly defining this step will make your proposed methodology more complete and reproducible.
Also, MediaPipe's visibility values should be described as visibility estimates, not as a guarantee of landmark accuracy. This distinction will improve the technical precision of your paper.

4.4 Handling Missing and Unreliable Landmarks :
Missing or unreliable body landmarks are common in surveillance videos, particularly when customers are partially hidden by shelves, products, or other people. These missing observations can affect the quality of pose sequences and make it difficult for the anomaly detection model to understand a person's actual movement.
To handle this problem, the proposed SentinelPose framework considers three methods for processing incomplete pose information.
Interpolation: When a body landmark is missing for a short period, its position can be estimated using valid observations from nearby frames. This helps maintain the continuity of movement patterns when the missing interval is small. However, interpolation is suitable only for short gaps because longer gaps may lead to inaccurate movement estimates.
Masking: A validity mask can be used to identify missing or unreliable landmarks. This helps distinguish unavailable observations from actual body coordinates. The mask can then be used during preprocessing or provided to a model designed to handle missing information.
Sequence Rejection: If a pose sequence contains too many missing or unreliable observations, it can be excluded from anomaly scoring. This prevents incomplete sequences from being treated as reliable representations of human movement.
The appropriate method will depend on the quality of the video data and the extent of missing information. The selected approach should be evaluated to determine its effect on anomaly detection performance.
To assess the overall quality of a pose sequence, SentinelPose proposes calculating a sequence-level reliability score using the frame reliability indicators:

$$Q_{\text{seq}} = \frac{1}{T}\sum_{t=1}^{T} R_t$$

where Trepresents the total number of frames in the sequence, and $R_t$indicates whether the pose in frame $t$satisfies the selected reliability threshold.
The sequence is considered suitable for anomaly analysis when its reliability score meets the minimum required threshold:

$$Q_{\text{seq}} \geq \beta$$

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Here, βrepresents the minimum proportion of reliable frames required for a sequence to be accepted. This threshold will be selected through validation experiments.
By applying this sequence-level reliability check, the proposed framework aims to reduce the influence of incomplete pose information on anomaly detection. It is also intended to reduce unnecessary alerts caused by unreliable observations. The actual effectiveness of this approach will be determined through experimental evaluation.

4.5 LSTM Autoencoder Architecture

The Long Short-Term Memory (LSTM) Autoencoder is proposed as the main model in SentinelPose for learning normal human movement patterns. LSTM networks are useful for analyzing sequential data because they can capture changes and relationships between observations over time.
The proposed Autoencoder consists of two main parts: an encoder and a decoder. The encoder takes a sequence of normalized human pose features and converts it into a compact representation of the movement. The decoder then uses this representation to reconstruct the original pose sequence.

In the initial design, each input sequence contains 30 consecutive video frames, with 132 pose features in each frame. Therefore, the input to the model is represented as:

$$X \in \mathbb{R}^{30\times132}$$

The encoder processes the input sequence and produces a compact representation:

$$h = f_{\text{enc}}(X)$$

where $h$represents the learned features of the movement sequence.

The decoder uses this representation to reconstruct the original input:

$$\hat{X} = f_{\text{dec}}(h)$$

where $\hat{X}$represents the reconstructed pose sequence.
The model is intended to be trained primarily on examples of normal shopping activities, such as walking through aisles, browsing shelves, reaching for products, and placing items into shopping baskets. By learning these movement patterns, the Autoencoder attempts to reproduce pose sequences similar to those observed during training.
During testing, the reconstructed sequence is compared with the original input to calculate a reconstruction error. Sequences that differ substantially from the learned patterns may produce higher reconstruction errors and can be considered potential behavioral anomalies. However, a high reconstruction error does not necessarily indicate suspicious behavior, as it may also result from unusual but harmless movements or inaccurate pose estimation.
The number of LSTM layers, hidden units, training epochs, and other model parameters will be determined during implementation and validation. The final configuration will be documented in the experimental setup to support reproducibility and allow the model's performance to be evaluated systematically.

4.6 Reconstruction Error Calculation

Reconstruction error is used to measure how closely the LSTM Autoencoder can reproduce an input pose sequence. In the SentinelPose framework, Mean Squared Error (MSE) is proposed to calculate the difference between the original pose features and the reconstructed features.
For an input sequence $X$and its reconstructed version $\hat{X}$, the reconstruction error is calculated as follows:

$$E(X) = \frac{1}{TF}\sum_{t=1}^{T}\sum_{j=1}^{F}(X_{t,j} - \hat{X}_{t,j})^2$$

where:

- $E(X)$represents the reconstruction error of the input sequence.
- Trepresents the number of frames in the sequence.
- Frepresents the number of features in each frame.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

- $X_{t,j}$represents the original value of feature $j$at frame t.
- $\hat{X}_{t,j}$represents the corresponding reconstructed value.

A lower reconstruction error indicates that the Autoencoder can reproduce the input sequence more accurately. In contrast, a higher error suggests that the observed movement differs from the patterns learned during training.

However, a high reconstruction error does not necessarily mean that a person is performing a suspicious activity. Unusual but harmless movements, incomplete pose information, and errors in landmark estimation may also increase the reconstruction error.

For this reason, SentinelPose uses reconstruction error as an indicator of potentially unusual movement rather than as direct evidence of wrongdoing. The proposed reliability and temporal verification mechanisms are then used to determine whether an observation should be considered for alert generation.

4.7 Anomaly Threshold Selection:

An anomaly threshold is needed to distinguish pose sequences that can be reconstructed relatively well from those that differ more strongly from the patterns learned during training.

Let τrepresent the anomaly threshold and $E(X_t)$represent the reconstruction error of the pose sequence ending at time t. The initial anomaly decision is defined as:

$$A_t = \begin{cases} 1, & E(X_t) > \tau \\ 0, & E(X_t) \leq \tau \end{cases}$$

Here, $A_t = 1$indicates that the reconstruction error exceeds the selected threshold. A value of $A_t = 0$indicates that the error does not exceed the threshold.

The threshold should be selected using validation data rather than the final testing dataset. One possible method is to calculate reconstruction errors for normal validation sequences and select a suitable percentile of the resulting error distribution as the threshold.

For example, a higher percentile produces a threshold that only larger reconstruction errors exceed. However, the appropriate percentile depends on the characteristics of the dataset and the desired balance between detecting unusual movements and limiting false alarms.

The performance of a single threshold may also vary across different camera angles, store layouts, and customer movement patterns. Therefore, this study will investigate how different threshold values affect anomaly detection performance and false alarm frequency. The final threshold will be determined through validation experiments and documented in the experimental setup.

4.8 Temporal Consistency Verification :

A high anomaly score in a single pose sequence may result from a temporary movement, a tracking interruption, or an error in pose estimation. Generating an alert immediately after such an observation may lead to unnecessary notifications.

To address this issue, SentinelPose proposes a temporal consistency verification mechanism. Instead of relying on a single anomaly decision, the system examines the anomaly indicators from several recent valid sequences to determine whether the unusual pattern continues over time.

Let K represent the number of recent valid sequence observations considered during verification. The temporal consistency score is calculated as:

$$C_t = \frac{1}{K} \sum_{j=0}^{K-1} A_{t-j}$$

where A_{t-j}represents the anomaly indicator for a recent valid sequence. The score $C_t$represents the proportion of the considered sequences whose reconstruction errors exceeded the selected anomaly threshold.

The temporal verification decision is defined as:

$$T_t = \begin{cases} 1, & C_t \geq \alpha \\ 0, & C_t < \alpha \end{cases}$$

Here, αrepresents the minimum proportion of anomalous observations required to satisfy the temporal condition.

For example, the system may require a specified proportion of recent valid sequences to exceed the anomaly threshold before generating an alert. The values of $K$and αwill be selected through validation experiments.

This approach is intended to reduce alerts caused by isolated high reconstruction errors. However, examining more observations before generating an alert may increase detection delay. The study will therefore evaluate both false alarm frequency and detection delay to understand the trade-off between alert reliability and response time.

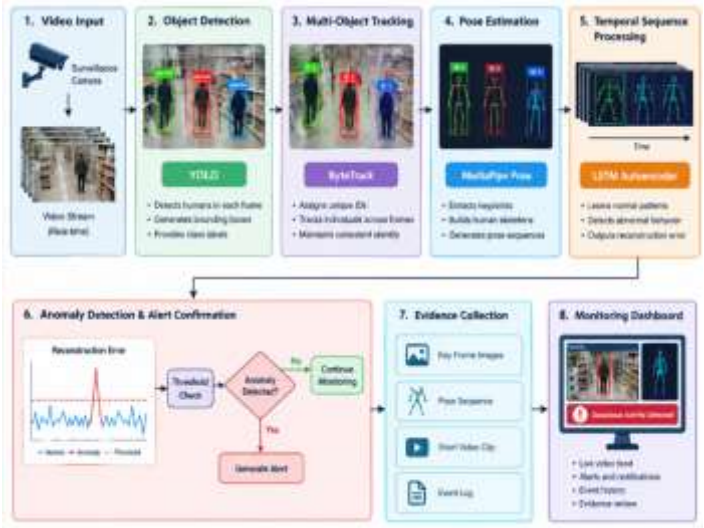
IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

4.9 Reliability-Aware Alert Decision Mechanism:



The proposed SentinelPose framework combines pose reliability assessment, reconstruction-based anomaly detection, and temporal consistency verification to determine when a potential anomaly should be reported.

The main objective is to ensure that an alert is generated only when the available pose information satisfies the reliability requirements and the observed movement meets the selected temporal anomaly conditions.

Let V_t indicate whether the pose sequence ending at time t satisfies the minimum reliability requirement. It is defined as:

$$V_t = \begin{cases} 1, & Q_{\text{seq},t} \geq \beta \\ 0, & Q_{\text{seq},t} < \beta \end{cases}$$

where $Q_{\text{seq},t}$ represents the reliability score of the current sequence and β represents the minimum acceptable proportion of reliable frames.

The final alert decision is expressed as:

$$D_t = V_t \cdot T_t$$

where:

- D_t represents the final alert decision.
- V_t indicates whether the current sequence satisfies the reliability requirement.
- T_t indicates whether the temporal verification condition is satisfied.

An alert is generated when $D_t = 1$, provided that sufficient valid observations are available and any required cooldown conditions are satisfied.

If a pose sequence does not meet the reliability requirement, it is treated as insufficient for reliable anomaly analysis. It is not automatically classified as normal behavior. This distinction is important because missing or unreliable information should not be interpreted as evidence that no unusual activity occurred.

The framework may also use an alert cooldown period to prevent repeated notifications for the same continuing event. When the alert conditions are satisfied, SentinelPose creates a potential-anomaly record and associates it with relevant video evidence, such as a saved video segment or a reference to the corresponding footage.

The resulting alert is presented to authorized security personnel for review. The system is intended to support human decision-making rather than automatically identify a person as a shoplifter or determine whether wrongdoing has occurred.

4.10 Proposed Algorithm and Mathematical Formulation

The proposed SentinelPose framework combines person detection, multi-object tracking, pose estimation, reliability assessment, temporal anomaly analysis, and alert generation into a single processing pipeline.

Algorithm 1 summarizes the main steps involved in analyzing a surveillance video and identifying potential behavioral anomalies.

Algorithm1: Reliability-Aware Temporal Behaviora -l Anomaly Detection:

Input: CCTV video stream or recorded video, trained LSTM Autoencoder, anomaly threshold τ, reliability thresholds γ and β, and temporal parameters K and α.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Output: Potential-anomaly alerts associated with relevant video evidence.

1. Read the next frame from the video source.
2. Detect people in the frame using YOLO.
3. Update person tracking information and temporary tracking IDs using ByteTrack.
4. Extract 33 body landmarks for each tracked person using MediaPipe Pose.
5. Normalize the extracted pose features and obtain the corresponding landmark visibility values.
6. Identify missing or unreliable pose observations.
7. Maintain a separate temporal pose buffer for each tracking ID.
8. Construct a fixed-length pose sequence when sufficient observations are available.
9. Calculate the sequence-level reliability score and determine whether the sequence satisfies the reliability requirement.
10. If the sequence is valid, calculate its reconstruction error using the trained LSTM Autoencoder.
11. Compare the reconstruction error with the selected anomaly threshold to obtain the initial anomaly indicator.
12. Update the temporal anomaly history using valid sequence observations.
13. Evaluate the temporal consistency score against the selected threshold α.
14. Generate a potential-anomaly alert when the reliability and temporal verification conditions are satisfied and the cooldown requirement permits an alert.
15. Store the alert information and associate it with the relevant video evidence.
16. Display the alert for review by authorized security personnel.
17. Continue processing until the video ends or the system is stopped.

The main decision stages of the proposed framework can be summarized as follows:

Pose Reliability Assessment
↓
Reconstruction Error Calculation
↓
Temporal Consistency Verification
↓
Reliability-Aware Alert Decision

This methodology provides a structured approach to investigating how pose quality and temporal movement analysis can be combined for retail surveillance. By considering the reliability of pose observations and examining anomaly scores over time, the proposed framework aims to reduce unnecessary alerts caused by incomplete observations and temporary movement variations.

The effectiveness of these mechanisms must be established through experiments using suitable datasets, baseline methods, and evaluation metrics. The implementation must also clearly define how missing observations, tracking interruptions, and insufficient temporal history are handled. These details are necessary to ensure that unreliable information is not incorrectly interpreted as normal behavior and that the proposed system can be evaluated consistently.

5. EXPERIMENTAL DESIGN AND EVALUATION METHODOLOGY

5.1 Dataset Selection and Description:

The selection of a suitable dataset is an important part of developing and evaluating the proposed SentinelPose framework. Since the system focuses on identifying unusual human movements in retail environments, the dataset should contain examples of normal shopping activities and unusual behaviors that may require further investigation.

The proposed study considers the PoseLift dataset, introduced for pose-based shoplifting detection in retail surveillance. This dataset is relevant because it focuses on human body movements in shopping environments and supports research on privacy-aware behavioral analysis.

The dataset selection process will examine the availability of pose landmarks, activity annotations, recording conditions, and permission for research use. The availability of original video footage will also be considered because the complete SentinelPose pipeline requires video input for person detection, tracking, and pose estimation.

If the selected dataset provides only pose sequences, it may be used to evaluate the anomaly detection and temporal verification modules. A separate suitable video dataset would then be required to test the complete video-processing pipeline.

The proposed experiments will focus on learning normal shopping movements and identifying sequences that differ from these patterns. The final dataset description will include the number of recordings, available activity categories, total duration, and training and testing samples after the selected data has been verified.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

5.2 Data Preprocessing and Dataset Splitting:

Before training the anomaly detection model, the selected data will undergo preprocessing to improve consistency and prepare it for temporal analysis.

For video-based data, frames will be extracted using OpenCV. YOLO will be used to detect people, while ByteTrack will maintain temporary tracking IDs. MediaPipe Pose will then extract body landmarks for each tracked individual.

Each pose observation will contain 33 body landmarks represented using horizontal position, vertical position, relative depth, and visibility information. This produces 132 features per frame.

The extracted landmark coordinates will be normalized to reduce the influence of person location and body size. Pose reliability assessment will also be applied to identify incomplete or low-visibility observations.

The processed features will be arranged into fixed-length temporal sequences. The initial design uses 30 frames per sequence, resulting in an input dimension of 30 × 132.

The dataset will be divided into training, validation, and testing subsets. Training data will mainly contain normal shopping activities so that the LSTM Autoencoder can learn common movement patterns. Validation data will be used to select model settings and anomaly thresholds. The testing subset will be kept separate for final performance evaluation and should contain suitable examples of normal and anomalous behavior.

To reduce data leakage, recordings or sessions will be assigned to separate subsets before overlapping temporal windows are created. This prevents nearly identical sequences from appearing in both training and testing data.

The final splitting proportions will depend on the amount of available data and the structure of the selected dataset.

5.3 Experimental Setup and Hardware Configuration:

The proposed SentinelPose framework will be developed using Python and commonly available computer vision and deep learning libraries.

OpenCV will be used for video processing, YOLO for person detection, ByteTrack for multi-person tracking, and MediaPipe Pose for body landmark extraction. The LSTM Autoencoder will be implemented using a suitable deep learning framework such as PyTorch or TensorFlow.

NumPy and Pandas will support numerical operations and data handling. Streamlit will be used to develop the monitoring dashboard, while SQLite will store alert information and evidence references.

The initial implementation is planned for a computer with 8 GB RAM and CPU-based processing, without requiring a dedicated GPU.

This hardware configuration is selected to investigate whether the proposed framework can operate on a relatively low-cost system. However, processing speed will depend on factors such as video resolution, the number of visible people, model size, and the computational requirements of pose estimation.

The experimental setup will record the operating system, processor details, Python version, library versions, input resolution, and model configuration to support reproducibility.

The final study will also examine whether the system can process recorded videos and live camera streams under the available hardware limitations.

5.4 Model Training and Validation Strategy

The LSTM Autoencoder in the proposed SentinelPose framework will be trained mainly on pose sequences representing normal shopping activities. The aim is to help the model learn common human movement patterns so that sequences that differ from these patterns may produce higher reconstruction errors.

Each training sample will contain a fixed-length sequence of normalized pose features. The encoder will learn a compact representation of the movement, while the decoder will use this representation to reconstruct the original pose sequence.

Mean Squared Error (MSE) will be used as the loss function during training. It measures the difference between the original pose features and the reconstructed features. The training loss is defined as:

$$L = \frac{1}{TF} \sum_{t=1}^{T} \sum_{j=1}^{F} (X_{t,j} - \hat{X}_{t,j})^2$$

where Trepresents the number of frames in a sequence, Frepresents the number of features in each frame, $X_{t,j}$represents the original feature value, and $\hat{X}_{t,j}$represents the corresponding reconstructed value.

During training, the model parameters will be adjusted to reduce the reconstruction loss on the training data. A separate validation dataset will be used to monitor the model's performance and select suitable hyperparameters. These may include the number of LSTM layers, the number of hidden units, batch size, learning rate, and number of training epochs.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Early stopping may also be used if the validation loss stops improving. This can help limit overfitting and prevent the model from learning the training data too closely at the expense of its ability to handle new observations.
In addition to the model parameters, the anomaly threshold, minimum pose reliability requirements, and temporal verification settings will be selected using validation data. These settings will determine how reconstruction errors are interpreted and when a potential anomaly should be reported.
Once the model configuration has been finalized, its performance will be evaluated on an independent test dataset. The test data will not be used to repeatedly adjust model parameters or decision thresholds. This separation helps provide a more reliable estimate of how the proposed framework may perform on previously unseen surveillance videos.

5.5 Baseline Models and Comparison Methods

To understand the contribution of the proposed reliability-aware temporal methodology, SentinelPose will be compared with simpler anomaly detection configurations.
The first baseline will use an LSTM Autoencoder with reconstruction-error thresholding. This configuration will generate anomaly decisions based on whether the reconstruction error exceeds a selected threshold, without applying the proposed reliability assessment or temporal verification mechanism.
A second configuration will combine LSTM reconstruction error with temporal consistency verification. This will help examine whether requiring persistent anomaly scores changes the frequency of unnecessary alerts.
A third configuration will combine the LSTM Autoencoder with pose reliability assessment. This will help investigate the effect of excluding or handling unreliable pose observations.
The complete proposed SentinelPose framework will combine both pose reliability assessment and temporal consistency verification.
All configurations will be evaluated using consistent datasets, preprocessing rules, and performance metrics. Where possible, the same trained LSTM Autoencoder will be used across configurations so that differences can be attributed more clearly to the proposed decision mechanisms.
These comparisons are intended to determine whether the complete framework offers measurable benefits over simpler reconstruction-based anomaly detection.

5.6 Ablation Study Design

An ablation study will be conducted to examine the contribution of individual components in the proposed framework.
The study will focus on two main additions: pose reliability assessment and temporal consistency verification.
Four experimental configurations are proposed:
Configuration A – Basic LSTM Autoencoder: This configuration will use pose sequences and reconstruction-error thresholding without additional reliability or temporal verification.
Configuration B – LSTM with Temporal Verification: This configuration will examine anomaly scores across multiple observations before generating an alert.
Configuration C – LSTM with Pose Reliability Assessment: This configuration will consider pose quality and exclude or handle unreliable observations before anomaly scoring.
Configuration D – Complete SentinelPose Framework: This configuration will combine the LSTM Autoencoder, pose reliability assessment, and temporal consistency verification.
The same evaluation procedure will be applied to all configurations.
The ablation study will compare detection precision, recall, F1-score, false alert frequency, and processing delay. It will also examine the number of observations rejected because of insufficient pose reliability.
This comparison will help identify whether each proposed component provides a useful contribution and whether combining the components introduces any trade-offs.
The results will be reported after implementation and testing. No improvement is assumed before the experiments are completed.

5.7 Performance Evaluation Metrics

The proposed framework will be evaluated using several performance metrics to assess its anomaly detection accuracy and suitability for practical retail surveillance.
Precision: Precision measures the proportion of detected anomalies that correspond to actual annotated anomalous events or observations.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Recall measures the proportion of annotated anomalies that are correctly detected by the system.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

$$Recall = \frac{TP}{TP + FN}$$

F1-Score: The F1-score provides a combined measure of precision and recall, helping evaluate the balance between correct detections and missed anomalies.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Here, TP represents true positives, FP represents false positives, and FN represents false negatives.
False Positive Rate: This metric measures the proportion of normal observations that are incorrectly classified as anomalous.

$$FPR = \frac{FP}{FP + TN}$$

Here, TN represents true negatives.
False Alerts per Hour: To assess performance in real surveillance settings, the number of false alerts will be measured relative to the total duration of the evaluated video.

$$FAH = \frac{N_{FA}}{H}$$

Here, N_{FA}represents the number of false alert events, and Hrepresents the evaluated video duration in hours.
Processing Speed: Processing speed will be measured in frames per second (FPS), indicating how many video frames the system can process per second.

$$FPS = \frac{N_f}{T_p}$$

Here, $N_f$represents the total number of processed frames, and $T_p$represents the total processing time in seconds.
The evaluation will distinguish between sequence-level anomaly classification and event-level alert generation. This distinction is important because multiple consecutive anomalous sequences may be associated with a single event. Event-level evaluation will therefore use a clearly defined matching rule to compare generated alerts with annotated anomalous events. The same rule will be applied consistently across all evaluated configurations.
Together, these metrics will help determine whether the proposed framework can detect meaningful behavioral anomalies, reduce unnecessary alerts, and maintain suitable processing performance for practical retail surveillance.

5.8 Real-Time Processing and Latency Evaluation :
Processing speed is an important factor in surveillance systems because delays in generating alerts can affect their usefulness in real-world situations. Therefore, the proposed SentinelPose framework will be evaluated in terms of processing time and alert generation delay.
The evaluation will measure the time taken by each major processing stage, including person detection, multi-person tracking, pose estimation, feature preprocessing, LSTM inference, and alert decision processing.
The approximate processing time per frame can be expressed as:

$$\boldsymbol{T_{\text{frame}} = T_{\text{det}} + T_{\text{track}} + T_{\text{pose}} + T_{\text{prep}} + T_{\text{infer}} + T_{\text{decision}}}$$

Here, each term represents the processing time required for its corresponding stage.
This equation provides a simplified estimate of processing time when the stages run sequentially. However, the actual processing time may vary if the implementation uses batching, parallel processing, or asynchronous video capture.
Alert latency will be measured as the time difference between the beginning of an annotated anomalous event and the generation of the corresponding alert.

$$\boldsymbol{L_{\text{alert}} = t_{\text{alert}} - t_{\text{event}}}$$

Here, t_{alert}represents the time when the alert is generated, while t_{event}represents the annotated start time of the event.
The study will investigate how sequence length, temporal verification settings, and pose reliability requirements influence alert latency. A longer verification period may help reduce unnecessary alerts, but it may also delay the reporting of

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

potential anomalies. Similarly, rejecting unreliable observations may improve the quality of the input data while reducing the amount of movement information available for analysis.

Experiments will be conducted using different video resolutions and numbers of tracked individuals, depending on the available test data. CPU utilization, memory usage, processing throughput, and alert latency will also be recorded to assess the computational requirements of the proposed framework.

The results will help determine whether SentinelPose can operate effectively on the selected hardware for practical surveillance applications. They will also help identify potential bottlenecks and areas where further optimization may be needed.

6. IMPLEMENTATION DESIGN

6.1 Software and Technology Stack

The proposed SentinelPose framework will be developed using Python because it provides suitable libraries for computer vision, deep learning, data processing, and application development. The system will combine several technologies to process surveillance videos, analyze human movements, and display potential-anomaly alerts.

OpenCV will be used for reading CCTV footage, extracting video frames, and performing basic image preprocessing. YOLO will identify people in each frame, while ByteTrack will track detected individuals across consecutive frames. MediaPipe Pose will extract body landmarks that represent the posture and movement of each tracked person.

The extracted pose information will be processed using NumPy and Pandas. An LSTM Autoencoder will be implemented using a deep learning framework such as PyTorch or TensorFlow. The model will learn normal movement patterns and calculate reconstruction errors to identify possible behavioral anomalies.

The proposed pose reliability assessment and temporal verification mechanisms will be implemented as additional processing components. These components will examine the quality of pose observations and the persistence of elevated anomaly scores before allowing the system to generate alerts.

Streamlit will be used to develop a monitoring dashboard through which authorized security personnel can view detected events and examine supporting evidence. SQLite will be used to store alert information, video source details, and review records.

The selected technologies are intended to support a modular implementation in which individual components can be tested and improved without changing the complete system.

6.2 System Module Integration

The proposed SentinelPose framework will be organized into separate software modules, with each module performing a specific task. This approach will make the system easier to develop, maintain, and evaluate.

The video acquisition module will receive footage from recorded videos or supported CCTV streams. OpenCV will extract frames and prepare them for processing.

The person detection module will use YOLO to identify individuals in each frame. These detections will be passed to ByteTrack, which will assign temporary tracking IDs and maintain the movement history of each person.

The pose estimation module will use MediaPipe Pose to extract body landmarks from the detected person regions. The extracted information will then be normalized and arranged into temporal sequences associated with the corresponding tracking IDs.

Before anomaly detection, the reliability assessment module will examine landmark visibility and incomplete observations. Sequences that do not satisfy the selected reliability requirements will be handled according to the defined preprocessing rules.

Valid sequences will be passed to the LSTM Autoencoder, which will calculate reconstruction errors. The temporal verification module will then examine whether elevated anomaly scores continue across multiple reliable observations.

When the defined alert conditions are satisfied, the alert management module will create a potential-anomaly record and save relevant evidence. The dashboard will display this information for human review.

Information will be exchanged between modules using defined data structures containing frame timestamps, tracking IDs, pose features, reliability scores, anomaly scores, and alert details.

The proposed modular design will also allow individual components to be evaluated separately during development. This will help identify processing delays, tracking problems, and errors that may affect the overall system.

6.3 Streamlit Monitoring Dashboard

A monitoring dashboard is proposed to provide a simple interface between SentinelPose and authorized security personnel. The dashboard will be developed using Streamlit, a Python-based framework for creating interactive applications.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

The main purpose of the dashboard will be to present potential-anomaly alerts in a clear and organized manner. Instead of requiring security personnel to examine every video frame continuously, the interface will highlight events that satisfy the proposed anomaly detection conditions.

The dashboard is planned to display the selected video source, processing status, generated alerts, anomaly scores, and supporting evidence.

Each alert will contain information such as the detection time, camera or video source, temporary tracking ID, anomaly score, and review status. Security personnel will be able to examine the associated image or video clip before making a decision.

The proposed interface will also provide filtering options to help users find alerts according to date, video source, anomaly score, or review status.

The review mechanism will allow authorized users to mark an alert as reviewed, dismissed, or requiring further investigation. These review decisions will be stored separately from the automatic anomaly predictions.

This distinction is important because SentinelPose is designed to identify unusual movement patterns rather than determine whether a person has committed shoplifting.

The dashboard will therefore act as a decision-support interface that helps security personnel examine potentially important events while keeping final judgments under human control.

6.4 Database and Evidence Storage

The proposed framework will use SQLite to store important information related to video processing, anomaly alerts, and user reviews. SQLite is selected because it is lightweight, does not require a separate database server, and is suitable for an initial local implementation.

The database will contain structured records for video sources, detected alerts, evidence references, review status, and model information.

Each alert record will be assigned a unique identifier and linked to details such as the timestamp, video source, temporary tracking ID, anomaly score, and current review status.

When an alert is generated, the evidence management module will save a relevant video frame or short video segment. The corresponding file path will be stored in the database so that the dashboard can retrieve the evidence when required.

The evidence files will be organized in separate storage directories. Images and video clips will be stored as files rather than directly inside database records to simplify file management.

The proposed system will also maintain model-related information, including the model version, training date, selected threshold, and sequence configuration. This information will help identify which model settings were used when an alert was generated.

Basic access controls and file-management rules will be considered to protect stored evidence and prevent unauthorized access.

Since surveillance footage may contain personal information, the implementation will aim to limit unnecessary data storage. Evidence retention periods and deletion procedures will need to follow the applicable privacy requirements and the policies of the organization using the system.

The database and evidence storage components are intended to support organized alert management and provide information for later system evaluation.

6.5 CPU-Based Deployment Considerations

The initial SentinelPose implementation is planned for a computer with approximately 8 GB RAM and CPU-based processing. This configuration is selected to investigate the possibility of running the proposed framework without relying on expensive graphics hardware.

However, processing multiple video frames through object detection, tracking, pose estimation, and an LSTM Autoencoder may require considerable computational resources.

To manage these requirements, the proposed implementation will consider lightweight detection models, controlled input resolution, and suitable frame-processing intervals.

The system may resize incoming video frames before detection to reduce processing time. A smaller YOLO model may also be selected to balance detection performance and computational requirements.

Pose estimation will be applied to the detected person regions, and temporal feature buffers will be maintained separately for each active tracking ID. The system will aim to avoid unnecessary repeated calculations and release temporary data when tracks are no longer active.

The LSTM Autoencoder will be designed with a manageable number of layers and hidden units so that inference can be performed using the available CPU resources.

The implementation will also consider separating video processing from dashboard updates. This may help prevent interface operations from unnecessarily delaying anomaly analysis.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Performance will be evaluated using frames per second, processing latency, CPU utilization, and memory consumption. The evaluation will also consider how processing requirements change when multiple people are visible in the same frame.

The proposed framework does not assume that the complete pipeline will achieve real-time performance on the selected hardware. Instead, experimental measurements will be used to determine its practical processing capacity and identify the optimizations required for deployment.

These considerations are intended to support the development of a low-cost and accessible surveillance framework while maintaining a clear balance between computational efficiency and anomaly detection quality.

6.1 Technology Stack:

Component	Proposed Technology	Purpose
Programming language	Python 3.x	Handles the main application logic
Video processing	OpenCV	Reads video frames, prepares them for analysis, and manages evidence
Person detection	Lightweight YOLO model	Detects people and identifies their bounding boxes
Tracking	ByteTrack	Tracks detected individuals across video frames
Pose estimation	MediaPipe Pose	Extracts key body landmarks for analysis
Sequence modeling	LSTM Autoencoder using PyTorch or TensorFlow	Learns temporal patterns and calculates anomaly scores
Numerical processing	NumPy	Performs feature calculations and sequence-related operations
Data processing	Pandas	Organizes datasets and manages evaluation results
Dashboard	Streamlit	Provides an interface for monitoring alerts and reviewing events
Database	SQLite	Stores alert details, evidence references, and configuration settings
Visualization	Plotly or Matplotlib	Displays anomaly scores, trends, and evaluation results

The final system should use either PyTorch or TensorFlow for the LSTM Autoencoder and clearly document the exact versions of all libraries and models used.

7. IMPLEMENTATION FRAMEWORK

7.1 Software and Technology Stack :

The proposed SentinelPose framework will be developed using Python because it provides suitable libraries for computer vision, deep learning, data processing, and application development. The system will combine several technologies to process surveillance videos, analyze human movements, and display potential-anomaly alerts.

OpenCV will be used for reading CCTV footage, extracting video frames, and performing basic image preprocessing. YOLO will identify people in each frame, while ByteTrack will track detected individuals across consecutive frames. MediaPipe Pose will extract body landmarks that represent the posture and movement of each tracked person.

The extracted pose information will be processed using NumPy and Pandas. An LSTM Autoencoder will be implemented using a deep learning framework such as PyTorch or TensorFlow. The model will learn normal movement patterns and calculate reconstruction errors to identify possible behavioral anomalies.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

The proposed pose reliability assessment and temporal verification mechanisms will be implemented as additional processing components. These components will examine the quality of pose observations and the persistence of elevated anomaly scores before allowing the system to generate alerts.

Streamlit will be used to develop a monitoring dashboard through which authorized security personnel can view detected events and examine supporting evidence. SQLite will be used to store alert information, video source details, and review records.

The selected technologies are intended to support a modular implementation in which individual components can be tested and improved without changing the complete system.

7.2 System Module Integration:

The proposed SentinelPose framework will be organized into separate software modules, with each module performing a specific task. This approach will make the system easier to develop, maintain, and evaluate.

The video acquisition module will receive footage from recorded videos or supported CCTV streams. OpenCV will extract frames and prepare them for processing.

The person detection module will use YOLO to identify individuals in each frame. These detections will be passed to ByteTrack, which will assign temporary tracking IDs and maintain the movement history of each person.

The pose estimation module will use MediaPipe Pose to extract body landmarks from the detected person regions. The extracted information will then be normalized and arranged into temporal sequences associated with the corresponding tracking IDs.

Before anomaly detection, the reliability assessment module will examine landmark visibility and incomplete observations. Sequences that do not satisfy the selected reliability requirements will be handled according to the defined preprocessing rules.

Valid sequences will be passed to the LSTM Autoencoder, which will calculate reconstruction errors. The temporal verification module will then examine whether elevated anomaly scores continue across multiple reliable observations.

When the defined alert conditions are satisfied, the alert management module will create a potential-anomaly record and save relevant evidence. The dashboard will display this information for human review.

Information will be exchanged between modules using defined data structures containing frame timestamps, tracking IDs, pose features, reliability scores, anomaly scores, and alert details.

The proposed modular design will also allow individual components to be evaluated separately during development. This will help identify processing delays, tracking problems, and errors that may affect the overall system.

7.3 Streamlit Monitoring Dashboard :

A monitoring dashboard is proposed to provide a simple interface between SentinelPose and authorized security personnel. The dashboard will be developed using Streamlit, a Python-based framework for creating interactive applications.

The main purpose of the dashboard will be to present potential-anomaly alerts in a clear and organized manner. Instead of requiring security personnel to examine every video frame continuously, the interface will highlight events that satisfy the proposed anomaly detection conditions.

The dashboard is planned to display the selected video source, processing status, generated alerts, anomaly scores, and supporting evidence.

Each alert will contain information such as the detection time, camera or video source, temporary tracking ID, anomaly score, and review status. Security personnel will be able to examine the associated image or video clip before making a decision.

The proposed interface will also provide filtering options to help users find alerts according to date, video source, anomaly score, or review status.

The review mechanism will allow authorized users to mark an alert as reviewed, dismissed, or requiring further investigation. These review decisions will be stored separately from the automatic anomaly predictions.

This distinction is important because SentinelPose is designed to identify unusual movement patterns rather than determine whether a person has committed shoplifting.

The dashboard will therefore act as a decision-support interface that helps security personnel examine potentially important events while keeping final judgments under human control.

7.4 Database and Evidence Storage :

The proposed framework will use SQLite to store important information related to video processing, anomaly alerts, and user reviews. SQLite is selected because it is lightweight, does not require a separate database server, and is suitable for an initial local implementation.

The database will contain structured records for video sources, detected alerts, evidence references, review status, and model information.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2026

DOI: 10.17148/IARJSET.2026.131065

Each alert record will be assigned a unique identifier and linked to details such as the timestamp, video source, temporary tracking ID, anomaly score, and current review status.
When an alert is generated, the evidence management module will save a relevant video frame or short video segment. The corresponding file path will be stored in the database so that the dashboard can retrieve the evidence when required. The evidence files will be organized in separate storage directories. Images and video clips will be stored as files rather than directly inside database records to simplify file management.
The proposed system will also maintain model-related information, including the model version, training date, selected threshold, and sequence configuration. This information will help identify which model settings were used when an alert was generated.
Basic access controls and file-management rules will be considered to protect stored evidence and prevent unauthorized access.
Since surveillance footage may contain personal information, the implementation will aim to limit unnecessary data storage. Evidence retention periods and deletion procedures will need to follow the applicable privacy requirements and the policies of the organization using the system.
The database and evidence storage components are intended to support organized alert management and provide information for later system evaluation.

7.5 CPU-Based Deployment Considerations :

The initial SentinelPose implementation is planned for a computer with approximately 8 GB RAM and CPU-based processing. This configuration is selected to investigate the possibility of running the proposed framework without relying on expensive graphics hardware.
However, processing multiple video frames through object detection, tracking, pose estimation, and an LSTM Autoencoder may require considerable computational resources.
To manage these requirements, the proposed implementation will consider lightweight detection models, controlled input resolution, and suitable frame-processing intervals.
The system may resize incoming video frames before detection to reduce processing time. A smaller YOLO model may also be selected to balance detection performance and computational requirements.
Pose estimation will be applied to the detected person regions, and temporal feature buffers will be maintained separately for each active tracking ID. The system will aim to avoid unnecessary repeated calculations and release temporary data when tracks are no longer active.
The LSTM Autoencoder will be designed with a manageable number of layers and hidden units so that inference can be performed using the available CPU resources.
The implementation will also consider separating video processing from dashboard updates. This may help prevent interface operations from unnecessarily delaying anomaly analysis.
Performance will be evaluated using frames per second, processing latency, CPU utilization, and memory consumption. The evaluation will also consider how processing requirements change when multiple people are visible in the same frame.
The proposed framework does not assume that the complete pipeline will achieve real-time performance on the selected hardware. Instead, experimental measurements will be used to determine its practical processing capacity and identify the optimizations required for deployment.
These considerations are intended to support the development of a low-cost and accessible surveillance framework while maintaining a clear balance between computational efficiency and anomaly detection quality.

8.CONCLUSION AND FUTURE WORK

8.1 Conclusion :

This paper presents SentinelPose, a proposed reliability-aware temporal framework for detecting unusual human behavior in retail surveillance environments. The main objective of the research is to support security personnel by identifying movement patterns that may require further investigation. The proposed framework focuses on human body movements rather than personal identity, making it suitable for exploring privacy-aware approaches to automated surveillance.
The system combines YOLO for person detection, ByteTrack for multi-person tracking, MediaPipe Pose for body landmark extraction, and an LSTM Autoencoder for analyzing movement patterns over time. The extracted pose landmarks are organized into temporal sequences, allowing the model to learn common shopping activities and identify possible behavioral anomalies through reconstruction error.
The main focus of this research is the proposed combination of pose reliability assessment and temporal consistency verification. Pose reliability assessment examines the quality of extracted body landmarks to identify incomplete or unreliable observations. Temporal consistency verification checks whether unusual movement patterns continue across

DOI: 10.17148/IARJSET.2026.131065

multiple valid sequences before an alert is generated. These mechanisms are intended to address problems caused by temporary pose estimation errors, tracking interruptions, and normal movements that may appear unusual.

The proposed framework also includes an alert management system that stores supporting video evidence and presents potential anomalies through a monitoring dashboard. Rather than automatically identifying a person as a shoplifter, SentinelPose is designed to provide information that security personnel can examine before making a decision.

An experimental methodology has been proposed to evaluate the framework using precision, recall, F1-score, false alerts per hour, processing speed, and alert latency. Baseline comparisons and ablation studies are planned to examine the individual contributions of pose reliability assessment and temporal verification.

Although the proposed approach aims to improve the dependability of behavioral anomaly alerts, its effectiveness has not yet been experimentally established. The planned evaluation will help determine whether the additional reliability and temporal mechanisms provide meaningful benefits under different surveillance conditions.

Overall, SentinelPose presents a structured approach to investigating reliable, pose-based behavioral anomaly detection for retail security. The research emphasizes the importance of considering input quality, movement consistency, computational limitations, and human review when designing automated surveillance systems.

8.2 Future Research Directions

The proposed SentinelPose framework can be further improved through the following research directions:

1. **Larger and Diverse Datasets:** Future research will focus on using larger retail surveillance datasets containing different shopping activities, camera angles, lighting conditions, and store environments to improve model generalization.

2. **Improved Pose Reliability Assessment:** More advanced methods can be developed to handle missing body landmarks, partial occlusion, and tracking errors for more reliable movement analysis.

3. **Adaptive Anomaly Thresholds:** Future improvements may include dynamic threshold selection based on movement patterns and environmental conditions to reduce unnecessary alerts.

4. **Advanced Deep Learning Models:** Alternative models such as GRU Autoencoders and Transformers can be explored and compared with the LSTM Autoencoder to improve temporal behavior analysis.

5. **Real-Time Performance Optimization:** Lightweight detection models and optimized processing techniques can be investigated to improve processing speed and reduce CPU and memory usage.

6. **Multi-Camera Surveillance Support:** The framework can be extended to analyze multiple CCTV streams and support monitoring across different areas of a retail store.

7. **Improved Alert and Evidence Management:** Future versions may include better alert filtering, event summaries, evidence organization, and feedback from security personnel.

8. **Privacy and Responsible Surveillance:** Further research will focus on protecting stored video evidence, maintaining user privacy, and ensuring that automated alerts remain subject to human review.

These future improvements aim to make SentinelPose more reliable, efficient, and suitable for practical retail surveillance applications

REFERENCES

[1]. N. Rashvand, G. A. Noghre, A. D. Pazho, S. Yao, and H. Tabkhi, “Exploring pose-based anomaly detection for retail security: A real-world shoplifting dataset and benchmark,” in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision Workshops (WACVW)*, 2025, doi: 10.1109/WACVW65960.2025.00125.

[2]. S. Yao, N. Rashvand, A. D. Pazho, and H. Tabkhi, “From offline to periodic adaptation for pose-based shoplifting detection in real-world retail security,” *IEEE Internet Things J.*, vol. 13, no. 8, pp. 16643–16652, 2026, doi: 10.1109/JIOT.2026.3660205.

[3]. N. Rashvand, G. A. Noghre, A. D. Pazho, B. R. Ardabili, and H. Tabkhi, “Shopformer: Transformer-based framework for detecting shoplifting via human pose,” in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 5798–5807, 2025.

[4]. O. Hirschorn and S. Avidan, “Normalizing flows for human pose anomaly detection,” in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 13545–13554, 2023, doi: 10.1109/ICCV51070.2023.01246.

DOI: 10.17148/IARJSET.2026.131065

[5]. H. Karim, K. Doshi, and Y. Yilmaz, "Real-time weakly supervised video anomaly detection," in *Proc. IEEE/CVF Winter Conf. Applications of Computer Vision (WACV)*, pp. 6848–6856, 2024, doi: 10.1109/WACV57701.2024.00670.

[6]. M. Zhang, J. Wang, Q. Qi, H. Sun, Z. Zhuang, P. Ren, R. Ma, and J. Liao, "Multi-scale video anomaly detection by multi-grained spatio-temporal representation learning," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, pp. 17385–17394, 2024, doi: 10.1109/CVPR52733.2024.01646.

[7]. E. Sengonul *et al.*, "Abnormal event detection in surveillance videos through LSTM auto-encoding and local minima assistance," *Discover Internet of Things*, vol. 5, Art. no. 32, 2025, doi: 10.1007/s43926-025-00127-3.

[8]. Y. Zhang *et al.*, "ByteTrack: Multi-object tracking by associating every detection box," in *Computer Vision – ECCV 2022*, Lecture Notes in Computer Science, vol. 13682, pp. 1–21, Springer, 2022, doi: 10.1007/978-3-031-20047-2_1.

© IARJSET This work is licensed under a Creative Commons Attribution 4.0 International License