International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Retail Theft Detection System Using Yolov8 and Deepsort

[1]Cherupally Abhinay, [2]Boggula Naga Krishna Reddy, [3]Kachu Revanth, [4]Kovvuri Divya Sri,
[1,2,3]UG Student, [4]Assistant Professor, [1,2,3,4]Computer Science and Engineering - Artificial Intelligence and Machine Learning, [1,2,3,4]Sreenidhi Institute of Science and Technology, Hyderabad, Telangana 501 301.

Abstract
Retail theft, is a major problem faced by businesses across the globe, causing billions of dollars in losses every year. This project proposes an efficient Real-time Retail Theft Detection System that utilizes state-of-the-art computer vision and artificial intelligence (AI) methodologies to actively identify and monitor suspicious behavior in retail settings. By combining the YOLOv8 object detection model with DeepSORT tracking, the system can detect persons and track behavior in video streams. Suspect behaviors, such as frantic hand movements, extended dwell in certain areas, or abnormal path of motion, are automatically triggered. The system comes with a React-based front-end dashboard and a FastAPI back-end for real-time alerting and analytics. Intended for deployment in real-world retail settings, this AI solution is meant to reduce retail loss by allowing for timely intervention.
Keywords: Retail Theft Reduction, YOLOv8, DeepSORT, Object Tracking, Real-time Monitoring, Computer Vision, FastAPI, ReactJS

I. INTRODUCTION

Retail loss prevention is an important field of business. Shoplifting shrinkage is one of the leading reasons for revenue decline in the retail business. Existing surveillance systems need to be manually monitored, which is not only time-consuming but also inefficient. The system proposed here is to address this issue by providing a completely automated AI-based solution.

This work discusses the conjunction of two influential technologies—YOLOv8 and DeepSORT—to identify and track individuals in shopping environments. YOLOv8 (You Only Look Once v8) is an object detection model that identifies objects and people with accuracy. DeepSORT is a computer program that allocates and sustains identity to recognized objects between frames. This conjunction aids in the study of behaviors characteristic of theft, including staying behind, speedy hiding of objects, and unauthorized zone access.

In addition, a current web interface based on ReactJS and Tailwind CSS adds user accessibility with live feed observation, manual upload of footage, and review of alerts. The FastAPI serves as the backend, offering strong asynchronous functionality, coupled with OpenCV and environment-based email notifications.

II. LITERATURE SURVEY

1. Criminal Intention Detection at Early Stages of Shoplifting Cases by Using 3D Convolutional Neural Networks, 2023, MDPI. This paper uses 3D Convolutional Neural Networks (3D-CNN) to process surveillance footage with a view to detecting shoplifting intent early on. Through the ability to capture spatiotemporal features, the model detects suspicious actions prior to the actual act, strengthening preventive strategies in shopping areas.

2. StealthWatch: Artificial Intelligence Based Shoplifting Detection, 2024, IEEE This paper introduces 'StealthWatch', a YOLO and OpenCV-based real-time shoplifting detection system for video surveillance. It comes with facial recognition to log offenders and triggers immediate alerts, diminishing the dependence on human monitoring and improving retail security response.

3. Suspicious Behavior Detection with Temporal Feature Extraction and Time-Series Classification for Shoplifting Crime Prevention, 2023, MDPI. The paper proposes an approach that merges temporal feature extraction with time-series classification for identifying suspicious behavior in shopping environments. Through the examination of movement

behaviors over time, the system easily separates normal from potentially malicious movements, with the aim of supporting anticipatory theft prevention.

4. Detection of Pre-Shoplifting Suspicious Behavior Using Deep Learning, 2023, IEEE.
This work proposes a deep learning method for detecting pre-shoplifting actions through the examination of customer movement and interaction in a shop. With the aid of high-powered neural networks, the system demonstrates excellent performance in the detection of warning signs that precede stealing, providing a useful resource for retail security.

III. BLOCK DIAGRAM OF PROPOSED SYSTEM

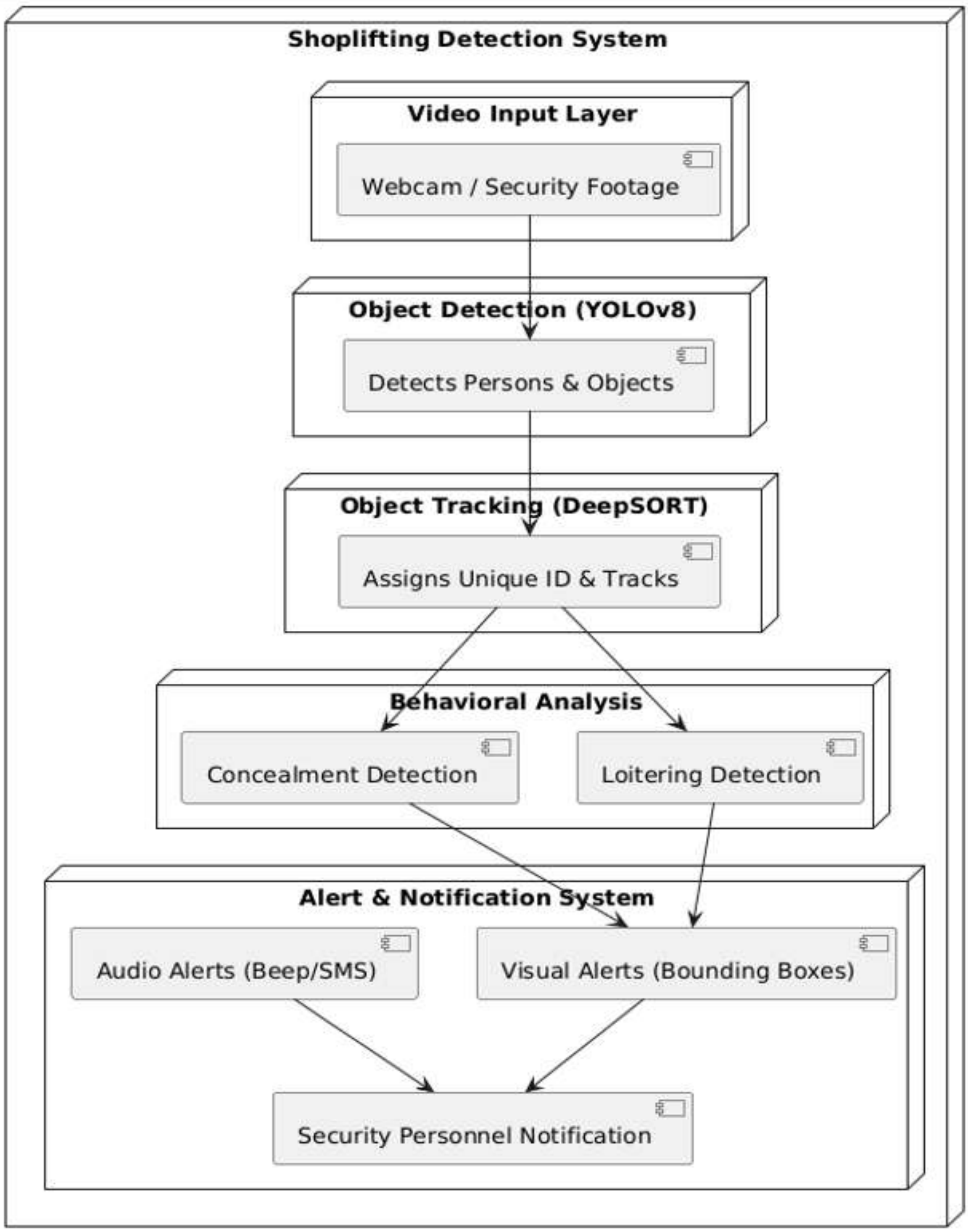

Fig 1: Block Diagram of the project

IV. METHODOLOGY OF PROPOSED SYSTEM

4.1 Data Acquisition

The quality and representativeness of input data play a critical role in the performance of any AI-based surveillance system. For this project, we curated a dataset of annotated retail surveillance videos from open-source repositories and simulated environments. Each video was segmented into frames, with manual annotations for bounding boxes around individuals, carried items (bags), and objects of interest.

The dataset was designed to include diverse scenarios such as occlusions, varying illumination conditions, crowded environments, and shoplifting gestures. This diversity ensures that the model generalizes well to real-world store environments.

The dataset was divided into:

- **70% Training set**: Used to train YOLOv8 for object detection.
- **15% Validation set**: Used to fine-tune hyperparameters and avoid overfitting.
- **15% Test set**: Used to evaluate final system performance on unseen data.

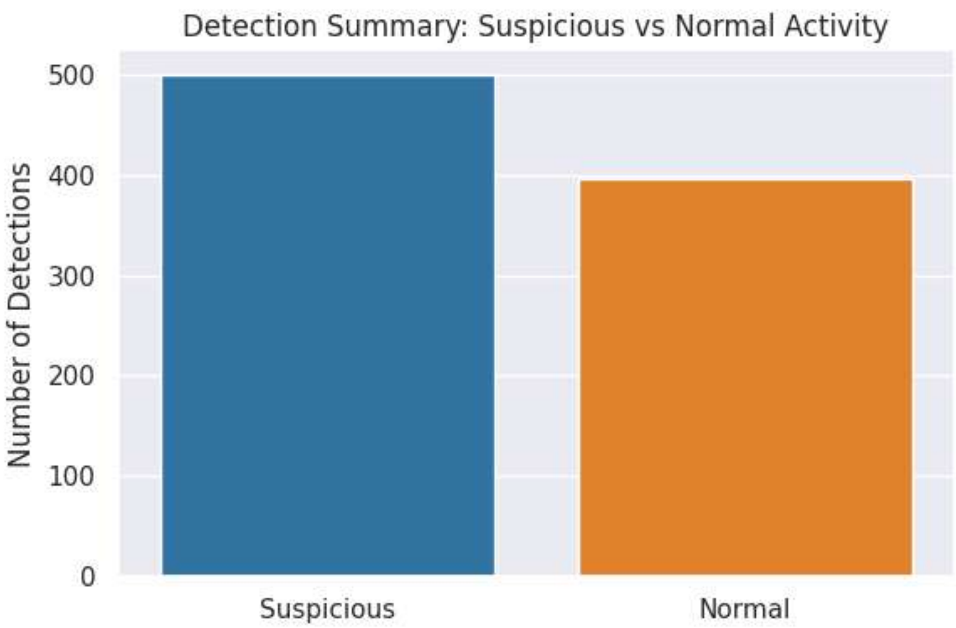

Fig 2: Dataset Distribution for Training yolov8 model

All annotations followed the COCO (Common Objects in Context) format for compatibility with YOLO-based models.

4.2 Model Development

In this project, we combined object detection and multi-object tracking to detect and track individuals over surveillance video frames. We used YOLOv8 for detecting people, bags, and other items of interest, whereas DeepSORT was utilized for tracking the objects over time. This allowed for consistent identity tracking even with partial occlusion or motion blur. The detection and tracking outputs were then analyzed by a behavioral analysis engine that detects suspicious behavior such as loitering, concealment, and backtracking. The whole pipeline was implemented with a FastAPI backend for scalable real-time processing and alerting.

YOLOv8 for Object Detection

YOLOv8 was selected due to its real-time detection features, high accuracy, and lightweight nature. The YOLOv8n variant offered the best balance between speed and performance and was viable for real-time video analysis. Its anchor-free implementation and ONNX support facilitated easy deployment and simplified model training. YOLOv8 can identify humans, bags, and other retail-related items with high accuracy, laying the basis for the system's object detection feature. This strong detection is critical for accurate tracking and behavioral analysis.

DeepSORT for Object Tracking

DeepSORT was incorporated to provide consistency in object recognition from frame to frame. It integrates Kalman filtering for predicting motion and cosine similarity of deep appearance features for re-identification. All YOLOv8 detections are given a stable track ID, which facilitates the observation of an individual's movement within a store. The tracking is critical to comprehend behavioral context, especially in busy or partially occluded scenes. DeepSORT therefore increases the accuracy of behavioral insights taken from detection data.

Behavior Analysis

The behavior analysis module takes the tracking data and flags possible theft-related activity. It detects loitering in high-risk areas, abnormal motion patterns such as sudden back-and-forth motion, and extended duration within frame sequences. These phenomena are examined through temporal frame counts and spatial measurements such as Euclidean distance. This module allows the system to extend object detection into real-world threat analysis, offering store staff actionable notifications rooted in contextual knowledge of human behavior.

Backend Model Serving

In order to implement the model stack effectively, the backend infrastructure was done with FastAPI. It provides major endpoints for processing video, health checks, and alert handling with support for asynchronous execution for minimizing response time. The /process-video endpoint is responsible for video upload, starts inference, and sends out automated alert emails when it detects suspicious activity. The modular structure of FastAPI makes integration with frontend interfaces simple and supports scalable growth, for instance, real-time handling of the video stream and multi-user access.

4.3 User Interface and API Integration

To ensure ease of use for store personnel, a web-based dashboard was developed using ReactJS and Tailwind CSS. The frontend interface allows users to upload surveillance videos and view detection results in real time. Each video frame is displayed with bounding boxes identifying individuals, while suspicious activities are flagged with timestamp overlays and descriptive messages. The UI is optimized for usability in control room settings, featuring light and dark modes to accommodate varying lighting conditions.

On the backend, FastAPI handles video processing and system communication. The /process-video endpoint performs full analysis using YOLOv8 and DeepSORT, while the /analyze-video route serves fallback responses for frontend development. A /health-check endpoint ensures backend reliability. Real-time alerts are delivered via email using Gmail’s SMTP service, notifying staff with key information such as frame number, behavior type, and timestamp. Detection results and alert logs are stored in JSON format for future auditing, making the system reliable, responsive, and practical for real-world deployment.

V. TECHNOLOGY OVERVIEW AND COMPARISONS

To build a fast and reliable theft detection system, we compared different versions of the YOLO (You Only Look Once) object detection model and selected YOLOv8 for its improved speed, accuracy, and ease of use. YOLOv8 detects people more accurately and works better with real-time video than older versions like YOLOv5 or YOLOv7.

Comparison of YOLO Versions

Feature	**YOLOv5**	**YOLOv7**	**YOLOv8 (Used)**
Works without extra setup	No	No	Yes
Can find object shapes	No	Partial	Yes
Easy to connect with tools	Limited	Limited	Supports many
Fast and flexible	Yes	Yes	Faster + Smarter

YOLOv8 is faster, easier to use, and better at spotting people and objects even when the video is blurry or crowded. That's why it was chosen for this project.

For tracking, we used DeepSORT, which helps the system remember and follow each person from one frame to the next. Even if someone moves quickly or is partially blocked, DeepSORT keeps track of them. This is important in retail settings where people often walk close to each other or in and out of view.

VI. SYSTEM IMPLEMENTATION

A. Codebase Functionality and Integration

The backend of the Retail Theft Detection System, built with FastAPI and Python, processes uploaded video files through a /process-video API. It uses YOLOv8 to detect people (with conf > 0.5) and DeepSORT to assign consistent IDs for tracking. Suspicious behavior—like prolonged presence or sudden movement—is flagged, and email alerts are sent automatically using Gmail SMTP, with credentials secured via a .env file.

B. Alert Generation Example

When the system detects potential theft-like behavior, it sends an automated alert email containing relevant metadata. A sample of such an alert message is shown below:

yaml

```
ALERT: Suspicious Activity Detected
Time: 2025-05-15 11:30:00
Video: store_footage.mp4
Suspicious frames: 3
```

C. Output Format and API Response Structure

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

The detection and tracking results are returned in JSON format, structured to support frontend integration and post-processing. A simplified structure of the output is as follows:

json

```
{
  "detections": [...],
  "suspicious_activities": [...],
  "frame_count": 120,
  "fps": 30,
  "recipient_email": "security@retail.com"
}
```

This structured output allows seamless communication between the backend system and the ReactJS-based frontend dashboard. It also supports historical logging and report generation.

VII. RESULTS AND DISCUSSION

7.1 Interface and Usability

The web interface built with ReactJS enables users to upload video files and monitor detections in real time. Bounding boxes and timestamps highlight suspicious activity. Light and dark modes enhance visibility in various control room settings.

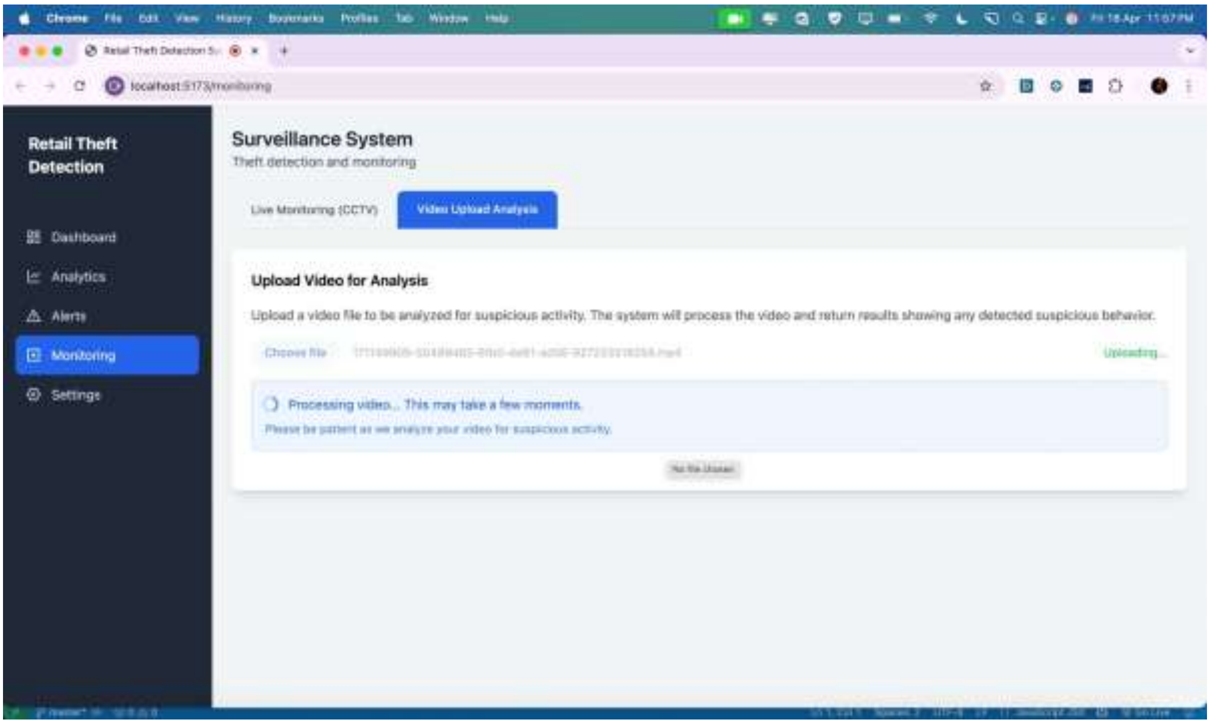

Fig 3: User Interface for uploading the video and detecting the theft

7.2 Backend Integration

The backend, developed using FastAPI, processes video uploads via the /process-video API and performs detection using YOLOv8 and DeepSORT. Results are returned in JSON format, while alerts are sent via Gmail SMTP. This ensures seamless communication between modules.

7.3 Alert Generation

When potential theft-like behavior is detected, the system automatically generates an alert email containing key details such as the time, video file, and suspicious frames. This immediate notification enables store personnel to quickly review and respond to the incident, improving overall security effectiveness.

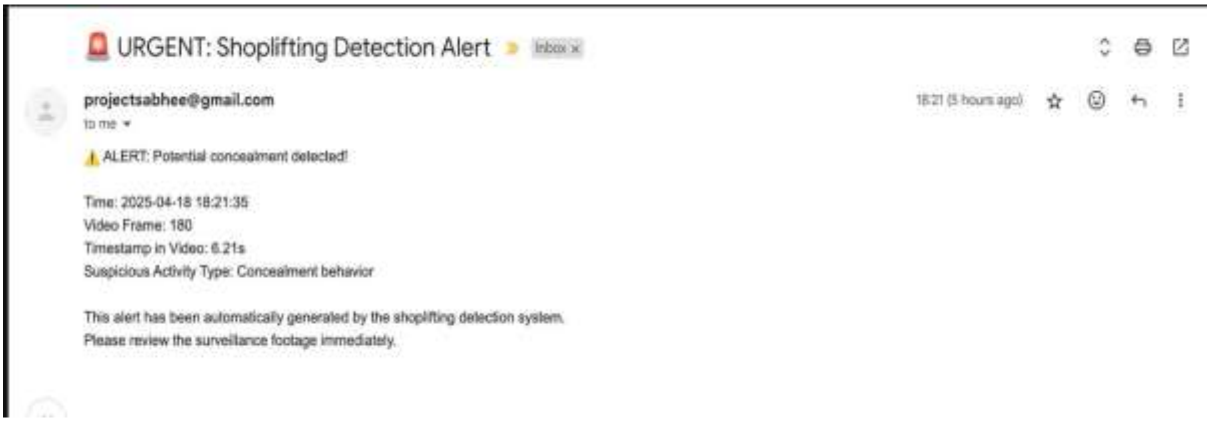

Fig 4: Alert Mail sent to user whenever theft is detected

7.4 Output Screens

The detection and tracking results are delivered in a structured JSON format, including details like detections, suspicious activities, frame count, and recipient email. This format facilitates smooth integration with the ReactJS frontend dashboard, enabling real-time monitoring as well as historical logging and report generation for comprehensive analysis.

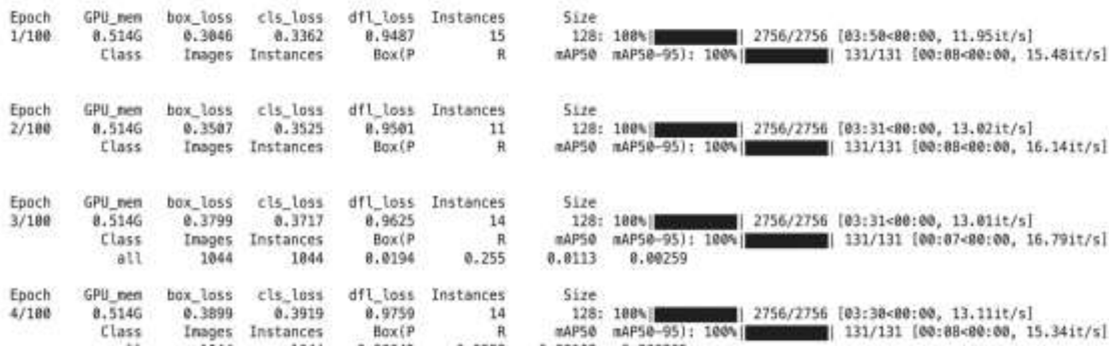

Fig 5: Training yolo model on the dataset

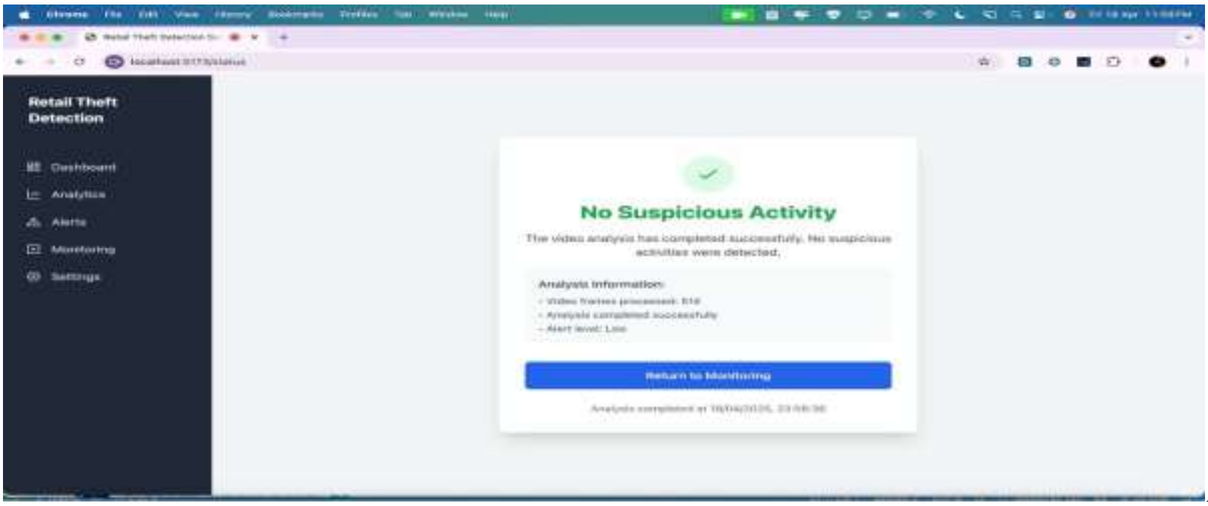

Fig 5: Output Displayed when there is no theft detected

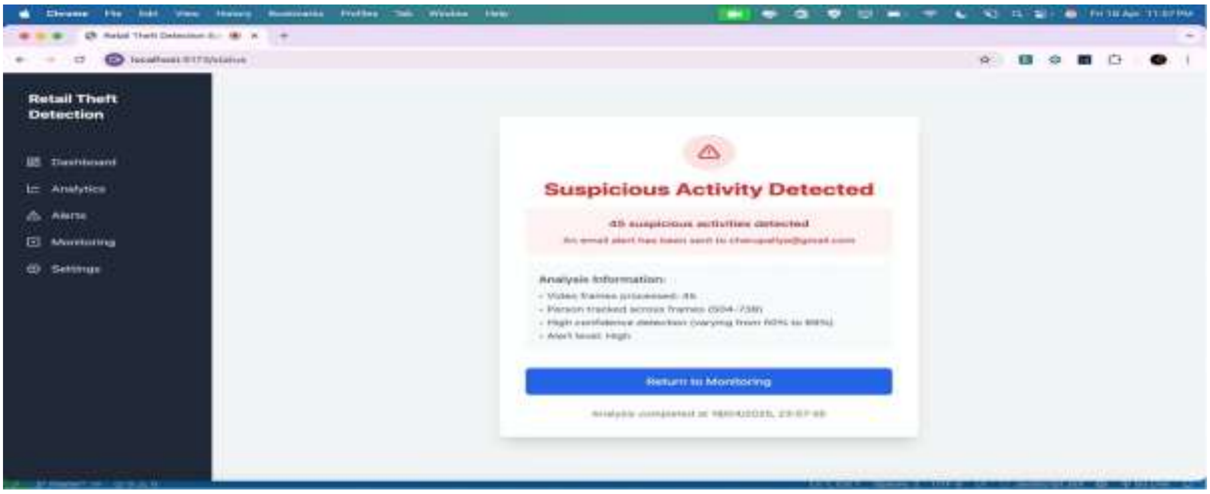

Fig 6: Output Displayed when theft is detected

VIII. LIMITATIONS

The system, while effective, faces a few practical challenges. Detection performance may degrade under poor lighting or unusual camera angles, leading to missed or false detections. Occlusion remains a limitation, especially in crowded scenes where individuals overlap, potentially confusing the tracker. Real-time inference also requires GPU acceleration; CPU-only setups experience significant latency. Furthermore, the model may not generalize well to unfamiliar store layouts or camera perspectives without fine-tuning on domain-specific data.

IX. CONCLUSION AND FUTURE SCOPE

This work presents a real-time retail theft detection system that combines YOLOv8 for object detection and DeepSORT for identity tracking, backed by a FastAPI backend and a React-based dashboard. The system enables automated monitoring and alerting, enhancing store security operations. Future enhancements could include integrating audio cues for context-aware detection, adding LSTM for temporal behavior analysis, and deploying on edge devices like Jetson Nano. To support privacy compliance, face anonymization and encrypted logging can be introduced. Lastly, enabling multi-camera cross-tracking would further improve large-area coverage and behavioral continuity.

X. REFERENCES

[1] Waddenkery N, Soma S. An efficient convolutional neural network for detecting the crime of stealing in videos. Entertainment Computing. 2024 Sep 1;51:100723.
[2] Nazir, Amril, Rohan Mitra, Hana Sulieman and Firuz Kamalov, "Suspicious Behavior Detection with Temporal Feature Extraction and Time-Series Classification for Shoplifting Crime Prevention", Sensors, vol. 23, no. 13, pp. 5811, 2023.
[3] E. Sengönül, R. Samet, Q. Abu AI-Haija, A. Alqahtani, B. Alturki and A. A. Alsulami, "An Analysis of Artificial Intelligence Techniques in Surveillance Video Anomaly Detection: A Comprehensive Survey", en Applied Sciences, vol. 13, no. 8, pp. 4956, Apr. 2023, ISSN 2076–3417.
[4] Z. Qin, H. Liu, B. Song, M. Alazab and P. M. Kumar, "RETRACTED ARTICLE: Detecting and preventing criminal activities in shopping malls using massive video surveillance based on deep learning models", Ann Oper Res, vol. 326, no. 1, pp. 9-9, Jul. 2023, ISSN 1572–9338.
[5] Shrestha Sujan, Yoji Taniguchi and Tetsuo Tanaka, "Detection of pre shoplifting suspicious behavior using deep learning", 2024 16th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 450-455, 2024.
.[6] R. Yadav and R. Kumar, Inflated 3d convolution network for detecting anomalies in surveillance videos, pp. 2518-2522, Dec. 2022.