IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

Analysing Human Behaviour in ATM Surveillance Footage

Sunitha Guruprasad
Associate Professor, Department of CSE
St Joseph Engineering College, Mangaluru, India

Abstract: *The ability of computers to predict potential crime scenes through the processing of CCTV footage can significantly enhance security. An intelligent automated system that can track, analyse, and efficiently detect irregular or suspicious activities in ATM environments would provide a proactive approach to crime prevention and surveillance. Researchers have proposed various approaches for crime prevention in ATM environments, including object detection methods using Fast R-CNN and Haar Cascade techniques for weapon detection, as well as pose-based abnormal behaviour detection models employing OpenPose, P3D ResNet, and CNN architectures for identifying anomalous activities. However, in most cases, surveillance cameras in ATM centres are still limited to mere recording, without real-time analysis or automated detection capabilities. The proposed method enhances ATM surveillance by integrating pose estimation, object detection, and human tracking techniques to detect abnormal activities. The system identifies individuals attempting to conceal their faces using objects such as helmets and promptly issues warnings, thereby promoting safer and more reliable ATM environments.*

Keywords: Surveillance, ATM, CCTV, Security, Crime

I. INTRODUCTION

In today's technologically advanced society, autonomous systems are becoming increasingly prevalent, and the banking sector has been significantly transformed by these innovations. One of the most impactful advancements is the Automated Teller Machine (ATM), which allows customers to perform a variety of financial transactions—such as fund transfers, cash withdrawals, bill payments, and account inquiries—beyond traditional banking hours. ATMs have thus become an essential component of modern banking, offering speed, convenience, and accessibility. However, as banks serve as critical points of financial activity, they are also prone to security risks and criminal incidents. With the continuous growth of society and rising living standards, the demand for secure and efficient financial services has increased, highlighting the need for advanced security technologies in ATM environments.

Automated Teller Machines (ATMs) have become an integral part of modern banking due to their convenience and efficiency. However, crimes occurring in self-service banking facilities remain a significant challenge. Traditional ATM surveillance systems primarily rely on manual monitoring and video recording, which serve mainly as post-incident evidence collection tools. This reactive approach often results in missed opportunities for timely intervention, and in many cases, the financial or personal losses incurred may be irreparable even when evidence is available.

With the rapid growth of automation and smart technologies, crimes targeting financial institutions have become increasingly prevalent. While CCTV cameras are widely deployed to enhance security, their reliance on continuous manual monitoring significantly limits their effectiveness. As surveillance networks expand, manual observation becomes both impractical and inefficient, underscoring the urgent need for intelligent and automated monitoring systems capable of real-time crime detection and prevention.

There is a growing need for advanced systems capable of detecting anomalies in ATM environments and reporting ongoing criminal activities to security authorities before suspect's escape. Criminal incidents, such as attacks involving weapons, can be identified as abnormal behaviours, thereby enhancing public safety, reducing crime rates, and preventing serious incidents. This research focuses on the detection of suspicious or criminal activities in ATMs, which

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

serve as critical banking facilities for secure financial transactions in public spaces. Although several studies have explored ATM crime detection, practical deployment remains limited due to inefficiencies and processing constraints in existing systems. Motivated by real-life incidents worldwide, the proposed system aims to address the rising challenges of ATM-related fraud and crime, including camera covering, money grabbing inside ATM centres, theft of ATM machines, and threatening behaviour. By leveraging intelligent surveillance, the system seeks to ensure safer transactions and strengthen the security of self-service banking.

An automated system is essential for efficiently tracking and detecting irregular or criminal activities within ATM environments. The proposed method leverages machine learning techniques—such as pose estimation, object detection, and human tracking—to enhance surveillance and security in ATMs. By utilizing ATM-installed cameras more effectively, the system aims to prevent attacks on machines and suspicious behaviours. For instance, individuals attempting to conceal their identity with helmets are detected and prompted to remove them. Additionally, pose-based abnormal activity detection is employed using the MediaPipe framework, which identifies 33 human body landmarks; by calculating angles between key landmarks and comparing them against predefined thresholds, abnormal behaviours can be recognized and flagged. To further improve monitoring, a centroid tracking algorithm is integrated to assign unique IDs to individuals, ensuring continuous tracking within the frame. If multiple persons are detected simultaneously in restricted ATM spaces, the system triggers a warning message. Together, these techniques contribute to a robust, automated, and intelligent ATM surveillance framework aimed at minimizing crime and ensuring safer financial transactions.

The proposed system integrates multiple machine learning techniques to enhance ATM surveillance and crime detection. Helmet detection is implemented using the YOLO (You Only Look Once) algorithm, a real-time object detection framework that formulates detection as a regression problem to simultaneously predict bounding boxes and class probabilities in a single forward pass through a convolutional neural network (CNN). A custom-trained YOLO model is employed to accurately identify individuals attempting to conceal their identity with helmets. In addition, the system incorporates MediaPipe Pose, a high-fidelity body pose estimation framework capable of inferring 33 three-dimensional landmarks from RGB video frames, enabling the recognition of abnormal human activities. To ensure continuous monitoring, a centroid tracking algorithm is used to assign unique IDs to detected individuals, facilitating real-time person tracking and counting. Together, these components provide a robust and intelligent surveillance mechanism for detecting suspicious behaviours in ATM environments.

II. LITERATURE SURVEY

Behaviour analysis of human actions in ATM surveillance systems has emerged as an effective approach for the automatic detection of abnormal activities. Conventional self-service banking surveillance largely depends on manual monitoring and post-event video review, where the primary objective is evidence collection after an incident occurs. However, such manual approaches are often inefficient, labour-intensive, and prone to delayed response. Recent research highlights the growing need for intelligent video analytics that can perform real-time detection and interpretation of suspicious behaviour, thereby enhancing both security and responsiveness in ATM environments.Shaik et al. [1] proposed a deep learning–based framework to enhance physical security in ATMs by leveraging intelligent video surveillance. Their workemphasizes the effective utilization of cameras inside ATM kiosks to reduce potential attacks on ATM machines. The system employs multiple deep learning algorithms to detect a customer's face, carried objects, and motion patterns upon entry. Furthermore, the framework is designed to identify suspicious behaviour such as attempts to conceal one's face with helmets, sunglasses, or other coverings. In such cases, the system proactively alerts the user by displaying a warning message, thereby aiming to mitigate risks before an incident occurs.

Sathis et al. [2] proposed a robust computer vision–based approach for the real-time identification of abnormal activities in public premises. Their system leverages digital image processing techniques combined with deep learning methods to detect anomalies in both images and video streams. The primary objective is to enable machines to achieve a human-like understanding of visual content by learning to recognize patterns and irregular behaviours. In their framework, whenever an abnormal event is detected, the captured image is automatically forwarded via email to the concerned authorities, thereby ensuring timely notification and response.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

Menglin et al. [3] developed an abnormal behaviour detection system that integrates human pose preprocessing with a P3D ResNet–based classification model. The framework employs the OpenPose algorithm to extract key points of human body posture from input images, which are then categorized into five classes: (1) no person, (2) single person, (3) possible following or pushing (two persons), (4) fighting (two persons), and (5) attacking the ATM machine (single person). The P3D ResNet classifier leverages this labelled data to identify abnormal behaviours with high accuracy, demonstrating the potential of combining pose estimation and deep learning for real-time ATM surveillance applications.

Yogameena et al. [4] proposed a Support Vector Machine (SVM)–based approach for human behaviour classification in crowds using projection and star skeletonization techniques. In their framework, background subtraction is first applied to separate moving objects from the static background. A blob detection subsystem is then employed to identify foreground pixels, which are subsequently grouped into blobs representing individuals or groups. These processed features are used for behaviour analysis and classification, thereby enabling more accurate detection of abnormal activities within crowded environments.Yangsen et al. [5] proposed a fall detection system that integrates real-time human pose estimation with Support Vector Machine (SVM) classification. The system leverages the center of gravity (CoG) of the human body as a key feature for fall identification. When the CoG is observed to be unstable or deviates significantly from the normal range, the system interprets this as a strong indication of a fall, thereby enabling timely detection of such critical events.Manjula et al. [6] proposed a system for the detection and recognition of abnormal behaviour patterns in surveillance videos using a Support Vector Machine (SVM) classifier. Their approach combines statistical features (SF) with Hu moments (HF) for feature description, which serve as the input to the SVM. By leveraging these handcrafted features, the system aims to effectively distinguish between normal and abnormal human activities, thereby enhancing the reliability of surveillance-based behaviour analysis.

Win et al. [7] proposed a feature-based human activity recognition framework using neural networks. Their approach employs background subtraction for foreground detection, where foreground intensities are separated from the static background. One noted challenge in this method is that certain body parts may appear partially detached from the main body in the detected image, which can affect the accuracy of recognition. Nonetheless, the framework demonstrates the potential of combining background modelling with neural network–based classification for effective human activity analysis.

Ahmed et al. [8] proposed a model for the automatic detection of student behaviours during group presentations, focusing on task-oriented analysis. The study emphasizes evaluating group presentations by examining the behaviour of individual members, thereby providing a more granular assessment of performance. Their framework is specifically designed to analyse video recordings of group presentation scenarios, highlighting the potential of behaviour recognition systems in educational and collaborative environments.

Thomas et al. [9] proposed a framework for detecting abnormal human behaviour using a video-based model. Since a video is essentially a sequence of images, their system first detects human objects in each frame and then extracts relevant features through feature extraction techniques. Recent advancements in video anomaly detection, including their work, highlight the effectiveness of deep learning architectures such as Convolutional Neural Networks (CNNs) combined with Long Short-Term Memory (LSTM) networks and Autoencoders (AEs). These hybrid models enable more robust feature representation and improve the accuracy of behaviour classification in dynamic video sequences.

Kyo et al. [10] conducted a study on CNN-based human behaviour recognition with channel state information (CSI). Their work utilized an open dataset comprising eight categories of human behaviours: bed, fall, pick up, run, stand up, sit down, walk, and empty. To enhance temporal feature representation, they integrated Long Short-Term Memory (LSTM) networks alongside CNNs. The proposed model achieved an impressive average accuracy of 94.59%, though with a significant training time of 61,882 seconds, highlighting both the effectiveness and computational cost of deep learning approaches in behaviour recognition tasks.Shih et al. [11] proposed a video-based abnormal human behaviour detection system aimed at psychiatric patient monitoring. Their approach employs a threshold-based Conditional Random Field (CRF) model to discriminate between normal and abnormal behaviours. By modelling temporal dependencies in patient activities, the system enhances the accuracy of abnormality detection, making it particularly useful in healthcare surveillance scenarios where continuous and reliable monitoring is essential.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

Although several studies have explored abnormal human behaviour detection using SVM, CNN, LSTM, and pose estimation techniques across domains such as healthcare, education, and public surveillance, most approaches remain domain-specific, computationally expensive, or limited to predefined activities. Existing methods often lack real-time adaptability, robustness in complex environments, and ATM-specific behaviour analysis such as forced withdrawals, loitering, or vandalism. This highlights the need for a generalized, efficient, and real-time framework tailored for ATM surveillance.

III. METHODOLOGY

The proposed work is implemented using three models' people counter model, helmet detection model and pose detection model.

Architecture of people counter model

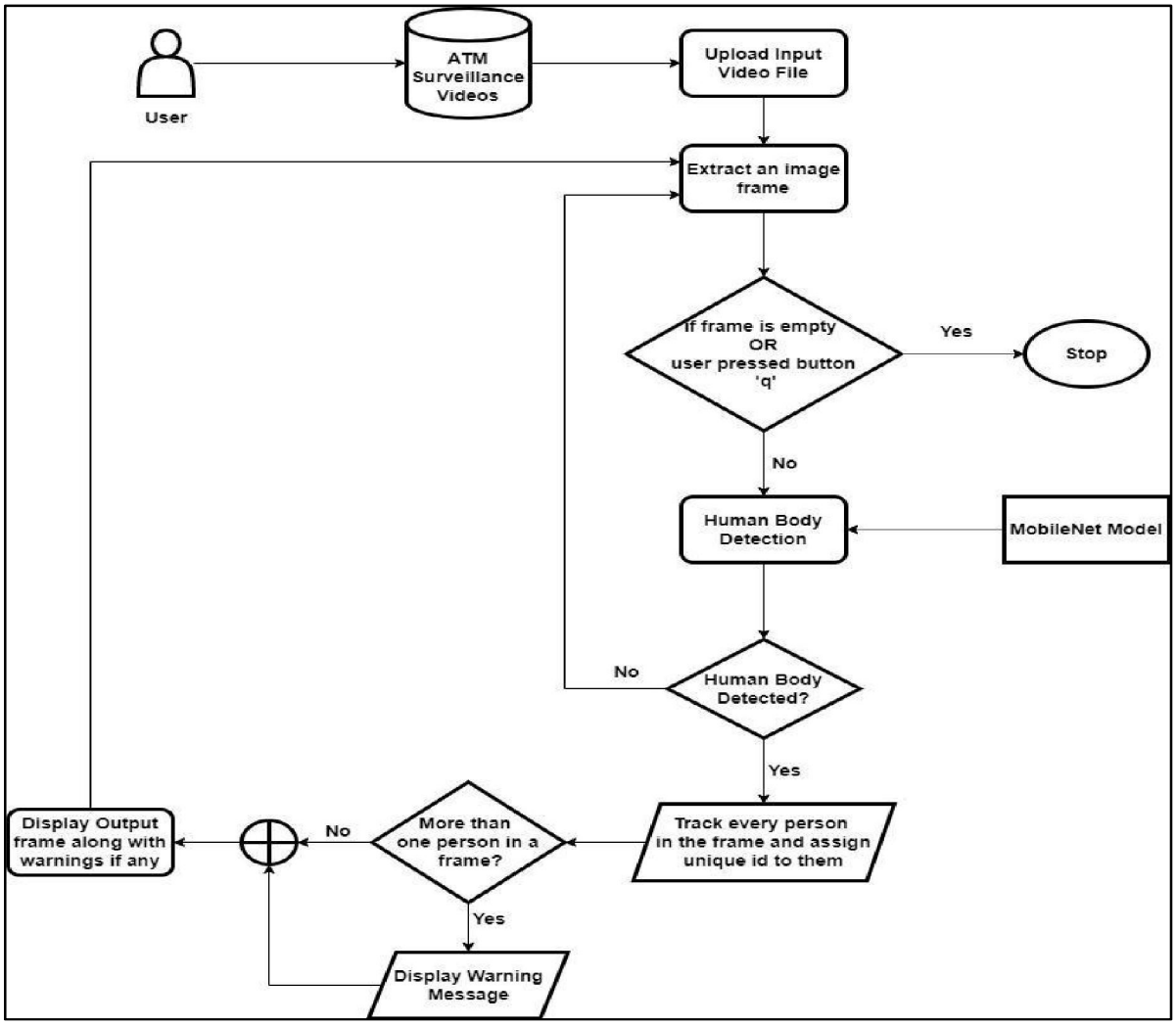

Fig. 1 Architecture of people counter model

In the proposed model, the user initially uploads a recorded video, which is then processed frame by frame. Within each frame, human objects are detected and their movements are tracked over time. To ensure consistent monitoring, a unique identifier (ID) is assigned to every detected individual, enabling the system to accurately determine the total number of people present in the scene. If the model detects more than one person simultaneously, a warning message is

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

generated to indicate potential abnormal activity. Conversely, if a frame contains no human presence, the system automatically terminates execution, thereby optimizing computational resources.Fig. 1 shows the architecture of people counter model.

Architecture of helmet detection model

In this model, the user begins by uploading a recorded video, which is processed frame by frame. For each frame, human objects and helmets are detected using the YOLOv3 model. The model is first trained on a dataset containing multiple annotated images of helmets to ensure accurate detection. During execution, if a person wearing a helmet is identified, the system generates a warning message to indicate abnormal behavior within the ATM environment. If a frame contains no human presence, the program automatically terminates execution, thereby conserving computational resources.Fig. 2 shows the architecture of helmet detection model.

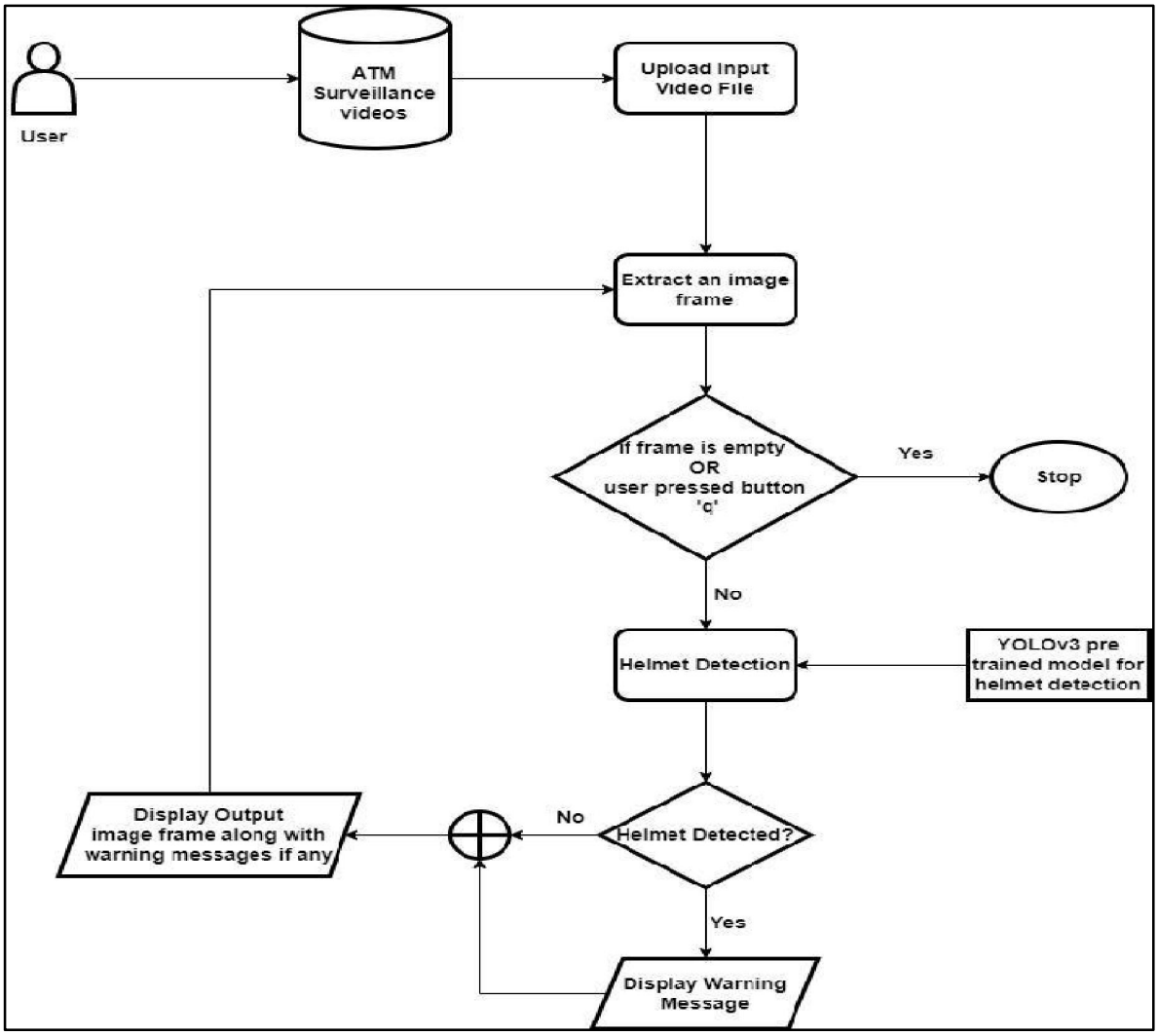

Fig. 2 Architecture of helmet detection model

Architecture of pose estimation model

In this model, the user first uploads a recorded video, which is then processed frame by frame. Within each frame, human objects are detected, and the MediaPipe pose estimation model is applied to identify 33 body landmarks. These

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

landmark coordinates are extracted, and the angles between specific joints are calculated to recognize abnormal activities such as bending, hand raising, or unusual postures. If no human is detected in a frame, the system automatically terminates execution, ensuring efficient resource utilization.Fig. 3 shows the architecture of pose estimation model

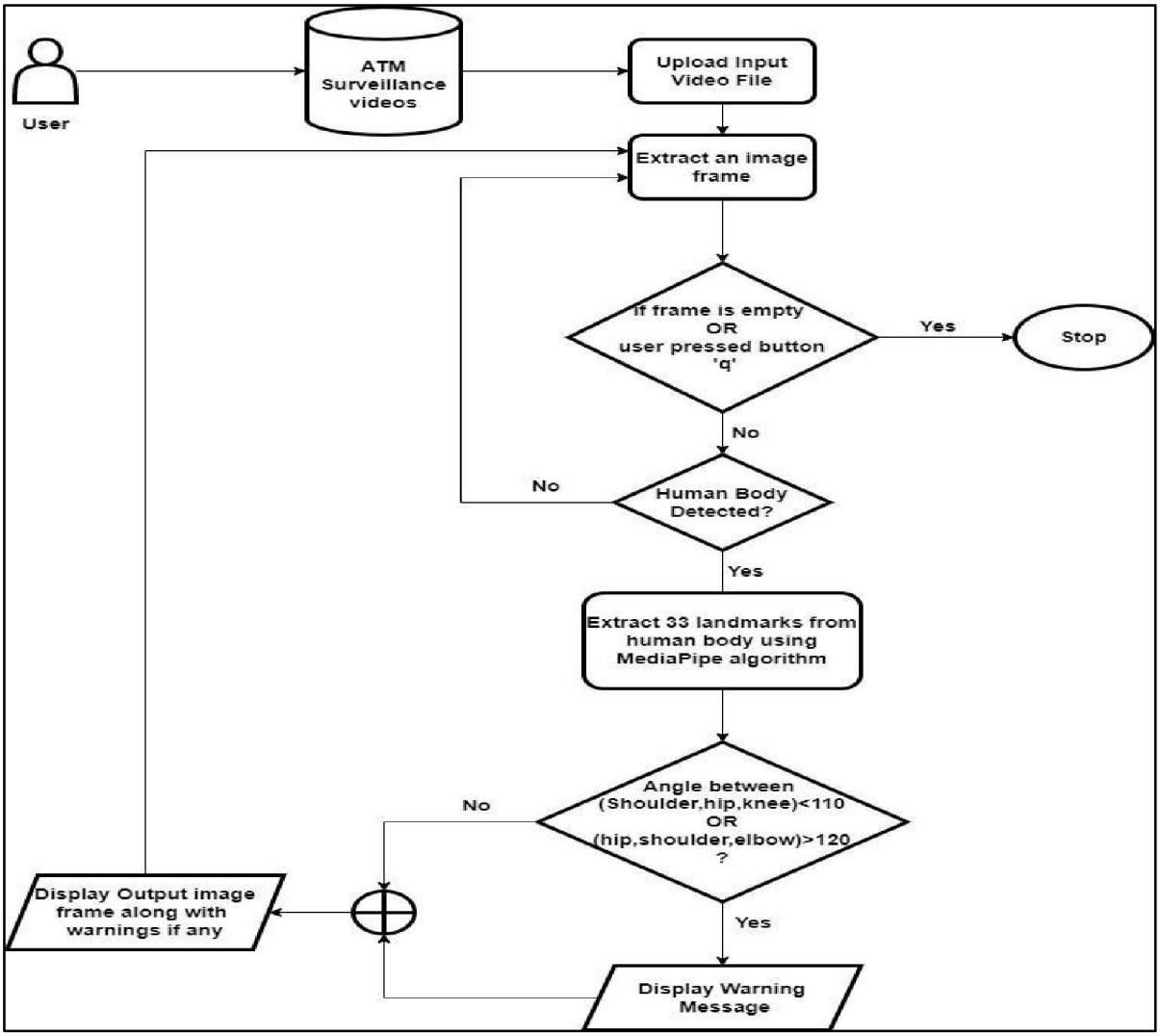

Fig. 3 Architecture of pose estimation model

IV. RESULTS AND DISCUSSION

Dataset and Implementation Setup

For this study, ATM surveillance video data was utilized as the primary dataset. A total of seven CCTV recordings, obtained from different surveillance systems, were considered for experimentation. The videos varied in duration, with an average length of approximately 1 minute and 30 seconds, thereby ensuring diversity in recording environments, illumination, and camera angles. Each sequence contained both normal and abnormal activities. Normal behaviours involved routine customer actions such as standing, withdrawing cash, or leaving the ATM premises, while abnormal behaviours included bending down, raising hands to obscure CCTV cameras, and wearing helmets inside ATM centers, which are typically considered suspicious.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

The videos were pre-processed frame-by-frame for human activity analysis. Human objects were detected using the YOLO framework, and 33 pose landmarks were extracted using the MediaPipe model. These landmarks were then used for behaviour classification based on joint angles and movement patterns. Frames without human presence were excluded to optimize computational resources. The dataset was manually annotated into normal and abnormal categories, with abnormal activities labelled according to their respective classes. This annotation process served as the ground truth for evaluating the performance of the proposed system.

The implementation was carried out using Python, with Python 3.7 recommended for compatibility. To deploy the proposed system as a web application, the Flask framework was employed due to its lightweight nature and ease of integration with machine learning models. Flask can be installed via the pip install flask command. Additionally, essential Python libraries such as OpenCV (cv2) for image processing, MediaPipe for pose estimation, NumPy for numerical computations, and OS for system-level operations were installed. These dependencies were obtained through the Python package manager (pip).

After environment setup and dependency installation, the system was capable of processing the dataset videos in real time. The combination of YOLO-based detection, MediaPipe landmark extraction, and Flask web deployment provided an efficient and scalable platform for abnormal activity detection in ATM surveillance footage.

Results of Helmet Detection Module

The helmet detection module was implemented using the YOLO (You Only Look Once) object detection algorithm. YOLO is a deep learning-based framework that employs convolutional neural networks (CNNs) to achieve real-time object detection. Among the various variants available (such as Tiny-YOLO and YOLOv3), the proposed system utilized YOLOv3, as it provides a balance between detection accuracy and computational efficiency.

When a helmet is detected within a video frame, YOLOv3 draws a bounding box around the object and triggers a warning message within the system. This enables real-time monitoring of abnormal activity, such as wearing helmets inside ATM premises, which is considered a suspicious behaviour.To evaluate the accuracy of the helmet detection module, a test video was used in which a person was wearing a helmet throughout the entire duration. The accuracy was computed using the following formula:

$$\text{Accuracy} = \frac{\text{Number of frames in which helmet is detected}}{\text{Total number of frames}} * 100$$

This approach ensures a straightforward yet effective evaluation of detection performance across video frames. Table I shows the accuracy of the module.

TABLE I: Accuracy of helmet detection module

Dataset	Number offrames	Number of times helmet detected	Number of times helmet not detected	Accuracy in percentage%
Dataset1	300	281	19	93
Dataset2	117	80	37	70

Results of person counter module

To count the total number of people in the video, an object tracking method was employed. Object tracking involves assigning a unique identity (ID) to each detected individual and maintaining this assignment consistently across successive frames. The process consists of the following steps:

- Initial Detection: An initial set of object detections is obtained, typically represented by bounding box coordinates around detected persons.
- ID Assignment: Each detected person is assigned a unique ID.
- Tracking Across Frames: As individuals move through the video frames, the system maintains their unique IDs, ensuring consistent tracking over time.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

This method enables accurate tracking of multiple people simultaneously and allows for the counting of unique individuals present in the video sequence. Object tracking is therefore a crucial component in building a person counter system.For performance evaluation, a video containing more than one person throughout the duration was used. The accuracy of the tracking module was calculated using the following formula

$$\text{Accuracy} = \frac{\text{Number of correctly tracked objects across frames}}{\text{Total number of objects across frames}} * 100$$

This ensures that both detection consistency and ID assignment stability are taken into account during evaluation. Table II shows the accuracy of the module.

TABLE III: Accuracy of person counter module

Dataset	Number of frames	Number of times tracked objects are detected	Number of times tracked objects are not detected	Accuracy in percentage%
Dataset3	89	66	23	75
Dataset4	117	110	7	94

Results of pose based abnormal behaviour detection module

For detecting pose-based abnormal human activities, the MediaPipe pose estimation framework was employed. MediaPipe identifies 33 body landmarks for each detected person, providing precise coordinates of key joints such as shoulders, elbows, wrists, hips, knees, and ankles. These landmarks were extracted frame-by-frame and analyzed to determine variations in human posture.

By leveraging these key points, the system was able to detect abnormal activities such as bending down, raising hands to cover CCTV cameras, and other suspicious gestures. The angle between specific landmark coordinates (e.g., hip–knee–ankle or shoulder–elbow–wrist) was computed to distinguish between normal and abnormal behaviours.

This landmark-based approach ensures robust pose estimation even in videos with varying illumination or partial occlusion, making it well-suited for real-world surveillance scenarios.

Fig. 4 shows the output of the three modules.

a. Output of helmet detection module

b. Output of people counter module

c. Output of pose estimation module

Fig. 4 Outputs of three modules

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 1, August 2025
Impact Factor: 7.67

V. CONCLUSION

In this work, we implemented a set of approaches to efficiently detect abnormal activities in ATM surveillance systems. The proposed model integrates machine learning techniques such as object detection, object tracking, and pose estimation, which collectively enhance the overall performance of the system. The framework successfully identifies abnormal scenarios, including a person wearing a helmet inside the ATM, multiple individuals attempting to withdraw money simultaneously, and abnormal body postures such as bending or raising hands to obscure CCTV cameras. In each case, the system generates real-time warning messages, enabling timely intervention and preventive action. This makes the model particularly valuable for ATMs that have suffered significant financial losses due to theft or vandalism.The experimental results demonstrate that the proposed algorithm can serve as an effective tool for automated surveillance and anomaly detection in security-critical environments. By leveraging real-time monitoring and intelligent analysis, the system reduces the reliance on manual supervision and improves overall situational awareness.

For future work, the system can be extended to detect a wider range of abnormal activities in ATMs. Machine learning models can be trained to recognize additional objects and behaviors, further strengthening the surveillance framework. Moreover, the system can be enhanced with an automated alert mechanism that sends warning notifications via email or SMS to the concerned authorities immediately upon detecting suspicious activity. Such extensions would make the system more robust, scalable, and practical for real-world deployment.

REFERENCES

[1]. D. Shaik and A. Sanjana, "ATM Fraud Detection Using Deep Learning," Journal of Tianjin University Science and Technology, vol. 54, no. 11, Nov. 2021.

[2]. N. R. Sathis Kumar and T. Nirmalraj, "Detection of Suspicious Activity in ATM Using Deep Learning," International Research Journal in Global Engineering and Sciences (IRJGES), vol. 5.

[3]. Wang, Menglin, et al. "Abnormal Behavior Detection of ATM Surveillance Videos Based on Pseudo-3D Residual Network." 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA). IEEE, 2019.

[4]. B. Yogameena, E. Komagal, M. Archana, and S. R. Abhaikumar, "Support Vector Machine based human behaviour classification in crowd through projection and star skeletonization," Journal of Computer Science.

[5]. Chen, Yangsen, et al. "Fall detection system based on real-time pose estimation and SVM." 2021 IEEE 2nd international conference on big data, artificial intelligence and internet of things engineering (ICBAIE). IEEE, 2021.

[6]. Manjula, S., and K. Lakshmi. "Detection and recognition of abnormal behaviour patterns in surveillance videos using SVM classifier." Proceedings of International Conference on Recent Trends in Computing, Communication & Networking Technologies (ICRTCCNT) 2019. 2019.

[7]. Oo, Win Myat, Bawin Aye, and Myo Min Hein. "Feature Based Human Activity Recognition using Neural Network." 2020 International Conference on Advanced Information Technologies (ICAIT). IEEE, 2020.

[8]. Fekry, Ahmed, Georgios Dafoulas, and Manal Ismail. "Automatic detection for students behaviors in a group presentation." 2019 14th International Conference on Computer Engineering and Systems (ICCES). IEEE, 2019.

[9]. Gatt, Thomas, Dylan Seychell, and Alexiei Dingli. "Detecting human abnormal behaviour through a video generated model." 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA). IEEE, 2019.

[10]. Hwang, Kyo-Min, and Sang-Chul Kim. "A study of cnn-based human behavior recognition with channel state information." 2021 International Conference on Information Networking (ICOIN). IEEE, 2021.

[11]. Hsu, Shih-Chung, et al. "A video-based abnormal human behavior detection for psychiatric patient monitoring." 2018 International Workshop on Advanced Image Technology (IWAIT). IEEE, 2018.