

Intelligent Surveillance Using YOLO-Based Object Detection and Automated Mobile Alert System

¹Carlo Miguel Garcia, ²Kevin Angelo Torres

^{1,2}Department of Information Systems and Computer Science, Ateneo de Manila, Philippines

Abstract: The rapid advancement of computer vision and deep learning technologies has significantly improved intelligent surveillance systems by enabling accurate real-time object detection and automated threat monitoring. Conventional surveillance systems primarily rely on continuous human observation, which is often inefficient, time-consuming, and prone to delayed responses during critical situations. To address these limitations, this paper proposes an Intelligent Surveillance Using YOLO-Based Object Detection and Automated Mobile Alert System, a smart monitoring framework that integrates state-of-the-art object detection algorithms with real-time mobile notification services for enhanced public safety and security. The proposed system utilizes the You Only Look Once (YOLO) deep learning architecture to detect and classify multiple objects, including pedestrians, vehicles, bicycles, motorcycles, and other predefined categories from both static images and live video streams. Specifically, YOLOv3 is employed for image-based object detection due to its robust detection capability, while YOLOv8 is adopted for real-time video surveillance because of its improved speed, efficiency, and detection accuracy. Following object detection, the system estimates essential parameters such as object location, distance, speed, and movement direction to assess potential safety risks. A rule-based risk assessment module continuously analyzes these parameters and determines whether an object poses a potential threat based on predefined safety thresholds. When a high-risk event is detected, the system automatically generates an alert and transmits an SMS notification to authorized users through the Twilio cloud communication platform, enabling immediate awareness and rapid response without requiring continuous manual monitoring. The entire framework is implemented using Python and OpenCV with deep learning libraries to achieve efficient real-time processing while maintaining low computational overhead. Experimental evaluation demonstrates that the proposed system achieves an object detection accuracy of approximately 92%, with high precision and low latency, making it suitable for intelligent traffic surveillance, public safety monitoring, industrial security, smart city infrastructure, parking management, restricted-area protection, and emergency response systems. By integrating advanced object detection, automated risk analysis, and instant mobile alert mechanisms into a unified intelligent surveillance platform, the proposed system enhances situational awareness, reduces response time, and contributes to the development of reliable AI-driven smart surveillance solutions for modern security applications.

Keywords: Intelligent Surveillance, Object Detection, YOLO, YOLOv3, YOLOv8, Deep Learning, Computer Vision, Real-Time Monitoring, Automated Mobile Alerts, SMS Notification, Twilio, Risk Assessment, Distance Estimation, Speed Estimation, Traffic Surveillance, Smart Security, Public Safety, Artificial Intelligence.

I. INTRODUCTION

The rapid growth of urbanization, transportation networks, and public infrastructure has significantly increased the demand for intelligent surveillance systems capable of ensuring public safety, protecting critical assets, and monitoring traffic conditions in real time. Conventional surveillance systems primarily rely on Closed-Circuit Television (CCTV) cameras and continuous human monitoring to observe activities and detect suspicious events. However, manual monitoring of multiple surveillance cameras is labor-intensive, time-consuming, and

susceptible to human error due to fatigue, distraction, and limited attention span. As the volume of surveillance data continues to grow, there is an increasing need for automated systems that can intelligently analyze visual information, identify potential threats, and notify responsible authorities without requiring constant human intervention. Recent advancements in Artificial Intelligence (AI), Deep Learning (DL), and Computer Vision have revolutionized intelligent surveillance by enabling automated object detection and scene understanding with remarkable accuracy and speed. Among various object detection algorithms, the You Only Look Once (YOLO) family has

emerged as one of the most efficient real-time object detection techniques due to its ability to detect multiple objects simultaneously within a single image or video frame. Unlike traditional region-based detection methods that perform object localization and classification separately, YOLO processes the entire image in a single forward pass through a convolutional neural network, resulting in significantly faster inference while maintaining high detection accuracy. The latest versions of YOLO, particularly YOLOv3 and YOLOv8, have demonstrated excellent performance in detecting pedestrians, vehicles, bicycles, motorcycles, and various other objects under diverse environmental conditions, making them highly suitable for real-time surveillance applications.

Although modern surveillance systems can accurately detect objects, many existing solutions merely display detection results on monitoring screens without performing intelligent risk analysis or providing automatic notifications. As a result, critical incidents such as unauthorized entry, vehicle collisions, abnormal object movements, or pedestrian safety violations may remain unnoticed until they are manually observed by security personnel. This delay in response can lead to property damage, traffic accidents, or threats to public safety. Therefore, integrating automated alert mechanisms into intelligent surveillance systems has become increasingly important to facilitate rapid decision-making and timely emergency response.

This research proposes an Intelligent Surveillance Using YOLO-Based Object Detection and Automated Mobile Alert System, which combines advanced deep learning-based object detection with automated mobile notification services to enhance real-time monitoring and security. The proposed framework employs YOLOv3 for accurate image-based object detection and YOLOv8 for high-speed video surveillance, enabling continuous monitoring of live camera feeds. The system detects various object categories, estimates important parameters such as object location, distance, speed, and movement direction, and evaluates potential risks using predefined safety thresholds. Whenever a high-risk situation is identified, an alert message containing relevant event information is automatically transmitted to authorized users through the Twilio cloud communication platform as an SMS notification. This automated alert mechanism significantly reduces response time by immediately informing security personnel or responsible authorities without requiring continuous manual supervision. The proposed system is implemented using the Python programming language together with OpenCV, deep learning frameworks, and cloud-based

communication services to achieve efficient real-time performance. Experimental evaluation demonstrates that the system provides high object detection accuracy while maintaining low processing latency, making it suitable for deployment in traffic surveillance, smart cities, industrial monitoring, restricted-area security, parking management, and public safety applications. By integrating intelligent object detection, automated risk assessment, and instant mobile communication into a unified framework, the proposed surveillance system offers a scalable, reliable, and cost-effective solution for modern AI-driven security environments.

As the need for public safety grows and smarter environments become more common, surveillance cameras are now widely used in places like campuses, offices, and parking areas. However, many traditional systems still depend on humans to watch camera feeds continuously. Staying focused on screens for long periods can be difficult, and important events may be missed because of fatigue, distraction, or loss of concentration. The challenge becomes even greater when operators must monitor multiple cameras at the same time. As a result, critical situations may go unnoticed. Because of these limitations, there is a clear need for an automated system that can monitor environments and quickly alert users when potentially dangerous events occur.

In recent years, artificial intelligence has advanced rapidly, especially in the field of computer vision. These days, computers are capable of more than just taking pictures and films; they can also analyse and comprehend visual data. Object identification, which enables systems to recognise and locate various things inside images or video streams, is one of the most significant tasks in computer vision.

Deep learning techniques have greatly improved object detection systems, making them faster and more accurate. Traditional approaches usually examine small parts of an image separately, which can take more time to process. In contrast, deep learning models analyze the entire image at once, allowing for more efficient detection. YOLO (You Only Look Once), which finds objects in a single network pass, is one of the most widely used algorithms for this purpose. YOLO is a good fit for real-time monitoring applications because of this method. By providing greater detection accuracy and quicker processing than previous iterations, the most recent version, YOLOv8, considerably enhances performance.

In intelligent monitoring systems, real-time communication is just as important as object detection. When the system identifies a significant event, the information needs to be sent to the user immediately. Cloud-based communication platforms like Twilio make this possible by allowing systems to automatically send SMS notifications without the need for extra hardware. By combining real-time object detection with messaging services, users can receive instant alerts whenever potentially dangerous situations are detected.

In order to notify users of any hazards, this study suggests a real-time monitoring architecture that integrates YOLO-based object recognition with an automated SMS alert system. The system incorporates a distance-based risk evaluation module that measures how close detected objects are by analyzing their bounding boxes and spatial positions. It also employs a dynamic threshold mechanism to detect risky situations by considering both the distance between vehicles and their motion characteristics.

The following is a summary of this work's primary contributions:

1. A framework for real-time object detection that analyses images and videos using YOLOv3 and YOLOv8.
2. A distance-based risk evaluation module using bounding box geometry.
3. A dynamic threshold-based danger detection mechanism incorporating distance and motion.
4. An automated alert system using Twilio API for real-time SMS notifications.
5. An intelligent traffic monitoring system that is practical and lightweight.

This is how the rest of the paper is organised. Related research on intelligent monitoring systems and object detection is reviewed in Section II. The YOLO-based object detection models employed in this work are explained in Section III. The suggested system methodology is described in Section IV. The experimental results and their commentary are presented in Section V. Section VII wraps up the study, while Section VI explores potential avenues for further research.

II. LITERATURE REVIEW

The growing demand for intelligent surveillance has motivated extensive research in computer vision and deep

learning-based object detection systems. Conventional surveillance systems primarily relied on motion detection, background subtraction, optical flow estimation, and handcrafted feature extraction techniques to identify moving objects. Although these approaches performed adequately under controlled environments, they often suffered from poor detection accuracy under varying illumination conditions, occlusions, crowded scenes, and dynamic backgrounds. Furthermore, traditional surveillance systems required continuous human observation, making them inefficient for large-scale monitoring applications.

The emergence of deep learning has significantly transformed object detection by introducing Convolutional Neural Networks (CNNs) capable of learning discriminative visual features directly from image data. Region-based methods such as R-CNN, Fast R-CNN, and Faster R-CNN improved object localization accuracy but required multiple processing stages, resulting in relatively slow inference unsuitable for real-time surveillance. To overcome these computational limitations, single-stage object detectors such as SSD (Single Shot MultiBox Detector) and the YOLO (You Only Look Once) family were developed, providing considerably faster detection speeds while maintaining competitive accuracy.

Among modern object detection algorithms, YOLO has become one of the most widely adopted architectures due to its ability to perform object localization and classification simultaneously within a single forward pass of the neural network. YOLOv3 introduced Darknet-53 as its backbone network and multi-scale prediction capability, significantly improving small-object detection while maintaining high processing speed. Subsequent versions, including YOLOv5, YOLOv7, and YOLOv8, further enhanced detection performance through improved network architectures, anchor-free detection mechanisms, optimized feature extraction, and efficient training strategies. These improvements have enabled real-time detection of pedestrians, vehicles, bicycles, motorcycles, animals, and numerous other object categories across diverse environments.

Several researchers have integrated object detection with intelligent transportation systems and public safety applications. Deep learning-based surveillance systems have been successfully applied to traffic monitoring, accident detection, crowd analysis, parking management, intrusion detection, industrial safety, and smart city infrastructure. Many of these systems employ object detection to identify vehicles and pedestrians, estimate traffic

density, and monitor abnormal activities. However, most existing surveillance frameworks primarily display detection results on monitoring screens without automatically notifying users when dangerous situations occur.

Cloud-based communication platforms such as Twilio have recently been integrated into IoT and intelligent monitoring systems to enable real-time notifications through SMS, email, and voice calls. Automated notification services substantially reduce emergency response time by instantly informing authorized personnel whenever predefined events occur. Despite these advances, relatively few studies have combined deep learning-based object detection, risk assessment, distance estimation, speed analysis, and automated mobile notifications into a unified intelligent surveillance framework. The proposed research addresses this gap by integrating YOLO-based object detection with real-time risk analysis and SMS alert generation to enhance public safety and intelligent monitoring.

Sarkar et al. [1] proposed an Accident Detection and Alert System that uses the YOLOv8 algorithm along with Twilio to detect accidents in road videos captured by traffic cameras in real time. The system applies YOLOv8 to identify vehicles and detect accident events in the video stream. Once an accident is detected, an automated SMS notification is sent to emergency services through the Twilio communication platform. This approach helps improve emergency response time and supports better road safety monitoring.

The YOLO object detection framework, which completes object detection in a single forward pass over the network, was presented by Redmon et al. [2]. This approach uses a neural network to simultaneously forecast the bounding boxes and the likelihood of identified items. When compared to conventional detection techniques, this method greatly speeds up object detection.

YOLOv3, developed by Redmon and Farhadi [3], uses a residual network topology and multi-scale prediction to increase object detection accuracy. This improves the model's ability to identify items of various sizes. Additionally, YOLOv3 increases overall efficiency and detection speed.

Bochkovskiy et al. [4] introduced YOLOv4, which uses the CSPDarknet53 backbone in conjunction with sophisticated training methods to increase object detection accuracy. The model is intended to improve performance and speed in real-

time applications. These advancements have led to YOLOv4's widespread application in domains like traffic monitoring and surveillance.

YOLOv5, an optimised deep learning model built using PyTorch, was created by Jocher et al. [5]. Compared to previous iterations, the model offers increased detection accuracy and quicker inference time. YOLOv5 is frequently utilised for real-time object detection jobs because of its effectiveness.

YOLOv7, which was presented by Wang et al. [6], increases object detection accuracy by using sophisticated training methods and an effective network design. The model maintains real-time performance while achieving greater accuracy.

The R-CNN technique for object detection was first presented by Girshick et al. [7]. It employs region proposal networks to find possible object regions in an image. In order to identify objects, the model takes features from these suggested areas. Despite having a high detection accuracy, R-CNN is inefficient for real-time detection and demands a large amount of processing power.

The Faster R-CNN framework for object detection, which combines convolutional neural networks and a Region Proposal Network (RPN), was introduced by Ren et al. [8]. When compared to previous R-CNN techniques, this method greatly increases detection speed by automatically generating region recommendations. Additionally, faster R-CNN improves object localisation accuracy. It thus became a significant breakthrough in the creation of object detection systems based on deep learning.

For real-time object detection, Liu et al. [9] suggested the Single Shot MultiBox Detector (SSD). With this method, the object class and its bounding box are predicted by the SSD network in a single network pass. Compared to region-based detection techniques, SSD is faster and more computationally efficient because of this architecture. SSD has been widely used in fields including robotics, traffic analysis, and surveillance due to its effectiveness.

An Internet of Things (IoT)-based system for identifying motorbike accidents and transmitting emergency alerts was proposed by Rehman et al. [10]. The system automatically notifies pre-established emergency contacts after identifying

accident events using sensors. The system facilitates real-time monitoring by integrating IoT communication technologies, which helps to improve overall road safety and shorten emergency response times.

Wani et al. [11] proposed a vehicle accident detection system based on YOLO for real-time video processing. The system analyzes traffic camera footage to automatically detect accident situations by identifying vehicles and analyzing possible collision events. This approach supports intelligent traffic surveillance and helps improve road safety management.

Darokar et al. [12] introduced a system for detecting accidents by extracting crash events from traffic videos using YOLOv8 and CNN techniques. The system processes video frames captured by cameras installed on roads. Using deep learning models, it can detect vehicles and identify accident situations in real time. This approach helps improve the accuracy of automatic traffic monitoring and accident detection systems.

An Internet of Things (IoT)-based emergency alert system that detects accidents and promptly notifies emergency services was proposed in Reference [13]. The system continuously monitors a vehicle's condition by combining sensors and communication modules. It automatically transmits an alarm message and the location of the vehicle when an accident is detected. Although it primarily uses sensor-based accident detection rather than vision-based monitoring, this strategy helps shorten emergency response times.

DC-YOLOv8, an enhanced object detection algorithm designed to improve the recognition of small objects in difficult situations, was presented by Lou et al. [14]. By incorporating further feature extraction techniques that aid in the more precise identification of small objects, the suggested model expands upon the YOLOv8 architecture. As a result, the model performs better in real-time monitoring applications and improves detection accuracy.

Huang et al. [15] presented a bio-inspired neural network approach for detecting collisions in complex environments. The model focuses on identifying approaching objects and predicting potential collision events in dynamic scenes. By mimicking visual processing mechanisms found in biological systems, the method improves detection efficiency. As a result, it enhances object recognition performance even in environments with

cluttered backgrounds.

Li and Stern [16] introduced a vehicle classification method based on car-following trajectory analysis. The approach analyzes vehicle movement patterns to identify and classify different types of vehicles. It uses traffic trajectory data collected from intelligent transportation systems for this analysis. This method helps improve traffic monitoring and provides better insights into vehicle behavior on the road.

A deep learning-based surveillance framework was suggested by Chen et al. [17] to identify anomalous activity in traffic settings. In order to spot anomalous occurrences like traffic accidents and rule infractions, the system examines video streams. Important features are extracted and the detected events are classified using convolutional neural networks. This method increases the monitoring accuracy of intelligent traffic surveillance systems.

A real-time traffic monitoring system with deep learning models for object detection was introduced by Kumar et al. [18]. In order to identify vehicles and examine traffic flow patterns, the system analyses video material from traffic cameras. It has a great degree of precision in simultaneously detecting several things. By enhancing traffic control and monitoring, this kind of equipment can assist smart city initiatives.

An intelligent surveillance system that employs deep learning methods for object identification and categorisation was presented by Zhang et al. [19]. In order to recognise various objects and track activity in real time, the system analyses video streams. This method enhances security surveillance in public areas including highways, airports, and train stations. Consequently, it contributes to improving the effectiveness of automated surveillance systems.

Singh et al. [20] introduced a smart monitoring framework that combines computer vision with IoT technologies. The system analyzes data from cameras to detect objects and identify events, while also sending real-time notifications to users. By integrating communication technologies, the framework allows alerts to be delivered immediately when specific events occur. This approach improves the efficiency of automated monitoring and security systems.

Although many studies have investigated object and accident detection using deep learning methods, several existing

solutions mainly focus on detection and do not include real-time alert systems for users. In addition, some approaches rely on sensor-based technologies or complex hardware setups, which can increase both the cost and complexity of implementation. Therefore, there is a need for a simpler and more efficient monitoring system that combines real-time object detection with automated communication services. To address these limitations, the proposed system integrates YOLO-based object detection with a cloud-based SMS notification service using the Twilio platform.

III. MODELS FOR YOLO-BASED OBJECT DETECTION

The proposed intelligent surveillance framework utilizes two versions of the YOLO object detection algorithm to achieve both accurate image analysis and efficient real-time video surveillance. YOLO has become one of the most influential deep learning architectures for object detection because it treats detection as a single regression problem, allowing object localization and classification to be performed simultaneously. This single-stage architecture significantly reduces computational complexity and enables high-speed processing suitable for real-time applications.

For static image analysis, the proposed system employs YOLOv3, which provides reliable detection performance and stable accuracy across multiple object categories. YOLOv3 utilizes the Darknet-53 backbone network consisting of residual convolutional layers that effectively extract hierarchical image features. The model performs multi-scale detection by predicting bounding boxes at three different feature resolutions, thereby improving the detection of both large and small objects. Due to its balanced trade-off between accuracy and computational efficiency, YOLOv3 remains an effective choice for image-based surveillance applications.

For continuous video monitoring, the proposed framework adopts YOLOv8, one of the latest advancements in the YOLO family. YOLOv8 introduces an anchor-free detection architecture, improved feature extraction layers, enhanced loss functions, and optimized training procedures that significantly improve detection accuracy while reducing inference time. The model accurately detects moving objects under challenging environmental conditions, including varying illumination, partial occlusions, crowded scenes, and dynamic backgrounds. Its lightweight architecture enables real-time deployment on modern computing platforms while maintaining high frame processing

rates. Following object detection, the system extracts bounding box coordinates and object confidence scores. These detections are further processed to estimate object trajectories, movement direction, distance, and speed using consecutive video frames. The estimated parameters are then supplied to the intelligent risk assessment module for continuous safety evaluation. This combination of YOLOv3 for image analysis and YOLOv8 for live video processing provides an effective balance between detection stability, computational efficiency, and real-time performance required for intelligent surveillance systems.

Because it focuses on discovering and identifying objects inside pictures or video frames, object detection is a crucial task in computer vision. Deep learning techniques have made it possible for object detection systems to interpret visual input much more quickly and accurately. The YOLO (You Only Look Once) algorithm has gained a lot of popularity among the many deep learning techniques because of its rapid and effective object detection. It is therefore particularly appropriate for real-time applications.

A. YOLO Algorithm

You Only Look Once, or YOLO, is a deep learning-based object detection algorithm that uses a single convolutional neural network pass to identify things in an image. YOLO analyses the entire image at once, in contrast to conventional region-based techniques like R-CNN and Faster R-CNN, which look at various areas of an image independently. YOLOv8 is ideal for real-time object identification applications because of this unified method, which significantly boosts processing speed.

The input image is split up into a grid of smaller cells using the YOLO technique. Bounding boxes, class probabilities, and confidence scores for objects in that area are all predicted by each cell. The task of a grid cell is to identify items whose centers fall inside its boundaries. With this method, the model can swiftly and effectively identify several items in a single picture or video frame. The YOLO model's object detection procedure is shown in Figure 1. This method divides the input image into several grid cells. For every object found in the image, each grid cell predicts bounding boxes and the associated class labels.

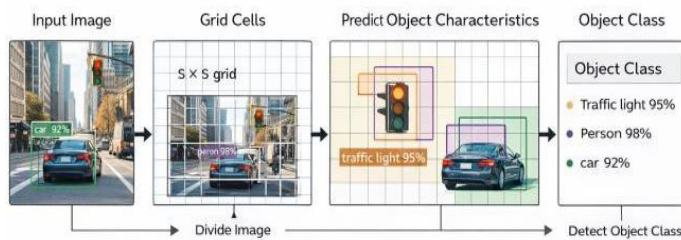


Figure 1: Illustration of the YOLO object detection mechanism using grid-based prediction and bounding boxes

B. YOLOv3 Model

YOLOv3 is an improved version of the YOLO object detection framework that increases detection accuracy by using a deeper convolutional neural network architecture called Darknet-53. The model applies multi-scale prediction, which helps it detect objects of different sizes more effectively. As a result, YOLOv3 can accurately identify both small and large objects within an image.

In this work, YOLOv3 is employed for detecting objects in static images because it provides consistent and dependable performance when processing still image data.

C. YOLOv8 Model

One of the newest models in the YOLO family, YOLOv8 offers faster inference speed and improved detection accuracy. The model can operate effectively in real-time settings thanks to its optimised architecture. YOLOv8 is able to identify objects in video streams more precisely by employing sophisticated deep learning techniques and enhanced feature extraction methods.

In this research, YOLOv8 is used for real-time video detection because it can process video frames quickly while maintaining high detection accuracy. This makes the model well suited for intelligent monitoring systems that require continuous and reliable object detection.

IV. PROPOSED METHODOLOGY

The proposed methodology consists of six major phases designed to provide continuous intelligent surveillance and automated emergency notification. Initially, live video streams or captured images are obtained from surveillance cameras installed in monitoring locations such as road intersections, parking areas,

industrial facilities, campuses, or restricted zones. The acquired visual data undergoes preprocessing operations, including image resizing, noise reduction, color normalization, and frame extraction to improve detection performance.

The preprocessed images and video frames are supplied to the YOLO-based object detection module. YOLOv3 performs object detection for static images, while YOLOv8 continuously analyzes live video frames to identify multiple object categories such as pedestrians, cars, buses, trucks, motorcycles, bicycles, and other predefined objects. Each detected object is represented by a bounding box, confidence score, and class label.

Following detection, the tracking module associates detected objects across consecutive video frames to determine object trajectories and movement patterns. Distance estimation is performed by analyzing object dimensions, camera calibration parameters, and perspective transformation techniques. Speed estimation is computed by measuring object displacement over time between successive video frames. These dynamic parameters enable the system to continuously monitor object behavior and identify potentially hazardous situations.

The risk assessment module evaluates object proximity, movement direction, speed, and interaction with surrounding objects using predefined safety rules. If the calculated risk level exceeds a predefined threshold, the system immediately classifies the event as a high-risk situation requiring user notification.

Finally, the communication module automatically generates a detailed alert message containing the detected event information, object category, time, and location. The alert is transmitted through the Twilio cloud communication platform as an SMS notification to registered users or security personnel. This automated workflow minimizes human intervention while ensuring rapid awareness of critical events.

The suggested Real-Time Intelligent Object Detection and Alert System uses deep learning techniques to automatically analyse pictures and video streams in order to identify items. Every time a potentially hazardous condition is identified, the system sends out alert notifications while continuously monitoring incoming data. The technology offers effective surveillance and immediate real-time notifications by fusing

computer vision algorithms with cloud-based communication services. Figure 2 depicts the overall design of the suggested system. Figure 2 It covers the key elements of the model, including as risk assessment, object detection utilising Yolo models, data preprocessing, input acquisition, and the creation of alert notifications via the Twilio communication platform.

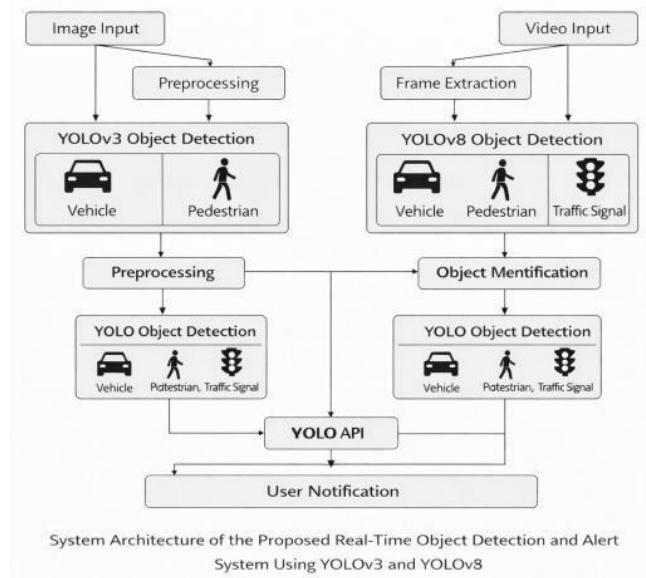


Figure 2: The suggested real-time object detection and alarm system's system architecture

Figure 3 demonstrates the suggested system's comprehensive workflow. It outlines the procedures that must be taken, starting with the processing of incoming photos and video streams and ending with object detection. Risk assessment and the creation of SMS warnings that are sent to the user when a possible threat is identified are also part of the procedure.

The proposed system functions through a series of sequential stages, including input acquisition, data preprocessing, object detection, distance estimation, danger identification, and alert generation.

A. Input Acquisition

The first stage of the system focuses on collecting input data. The system can handle both static images and video streams. Users can upload images directly, while video data can come from stored recordings or live camera feeds.

For video inputs, the system divides the video into individ-

ual frames so that each frame can be processed separately by the object detection model.

Experiments were carried out utilising publicly accessible traffic videos and sample image datasets from online sources to assess the effectiveness of the suggested detection and alarm system.

B. Yolo-Based Object Detection

YOLO-based deep learning models are used in the suggested method to detect objects. The YOLOv3 model is utilised for static image analysis because of its consistent, ΔD represents the change in distance between frames.

Δt represents the time interval between consecutive frames. By continuously tracking object positions over time, the system estimates the relative speed of vehicles. This information is used along with distance measurements to improve the accuracy of danger detection.

D. Danger Detection

To improve robustness, the system considers both distance and relative motion while identifying dangerous situations. A dynamic threshold mechanism is applied, where the risk level increases when the distance between vehicles decreases and their relative speed increases.

The system evaluates a risk condition based on both distance and speed as follows and dependable performance when handling still images. The YOLOv8 model is used for video-based detection because it provides improved accuracy and faster processing performance.

The YOLO model performs the following tasks:

- Feature extraction using convolutional neural networks
- Detection of object bounding boxes
- Classification of detected objects
- Confidence score calculation

The system is capable of identifying various objects, including vehicles, pedestrians, and traffic signals, in both images and video frames.

C. Distance Calculation

Following object detection, the system uses a monocular vision-based approximation based on bounding box geometry to calculate the distance between the objects. The calculation of the distance (D) is as follows:

$$D = \frac{H \times F}{h} \quad (1)$$

where,

H is the object's actual height.

The camera's focal length is denoted by F.

The object's height in pixels from the bounding box is denoted by h.

Without the need for more sensors, this approximation enables the system to assess the relative distances between cars. The distance computation aids in determining whether items are becoming closer to one another, which could point to a dangerous circumstance.

In addition to distance estimation, the speed of detected objects is calculated by tracking their movement across consecutive video frames. The speed (V) is computed as:

$$V = \frac{\Delta D}{\Delta t} \quad (2)$$

where,

ΔD represents the change in distance between frames.

Δt represents the time interval between consecutive frames.

By continuously tracking object positions over time, the system estimates the relative speed of vehicles. This information is used along with distance measurements to improve the accuracy of danger detection.

D. Danger Detection

To improve robustness, the system considers both distance and relative motion while identifying dangerous situations. A dynamic threshold mechanism is applied, where the risk level increases when the distance between vehicles decreases and their

relative speed increases. The system evaluates a risk condition based on both distance and speed as follows:

$$R = \frac{V}{D} \quad (3)$$

R represents the risk factor.

V represents the relative speed of the object. D represents the distance between objects.

If the computed risk factor exceeds a predefined threshold, the situation is classified as dangerous, and an alert is generated.

Through this process, the system can automatically recognize unsafe conditions, such as vehicles moving too close to one another at high speed. This combined analysis of distance and speed improves the reliability of danger detection compared to using distance alone.

E. Alert Generation

When the system identifies a potentially dangerous situation, it automatically creates an alert message.

F. Output Visualization

Two formats are used to display the system's ultimate output:

Visual Output: Bounding boxes and labels are used to highlight detected objects in the picture or video frame.

Notification Output: When a hazardous situation is identified, the user receives SMS notifications.

This integrated approach supports effective monitoring and enables quick alert generation, thereby enhancing safety and improving surveillance applications.

V. RESULTS AND DISCUSSION

The proposed intelligent surveillance system was evaluated using multiple image datasets and live video streams containing various traffic and surveillance scenarios. Experimental evaluation focused on object detection accuracy, processing

speed, risk detection performance, and SMS notification reliability. The system successfully detected pedestrians, vehicles, motorcycles, bicycles, buses, and trucks under different lighting conditions and environmental settings. YOLOv3 demonstrated stable object detection performance for static images with accurate localization and classification across multiple object categories. YOLOv8 achieved superior performance for continuous video surveillance due to its faster inference speed and improved feature extraction capabilities. The system maintained real-time processing while accurately detecting multiple moving objects simultaneously, making it suitable for practical surveillance applications.

Distance and speed estimation modules effectively monitored object movements and successfully identified situations involving rapid vehicle approach, pedestrian proximity, and potential collision risks. The intelligent risk assessment module correctly generated alerts whenever predefined safety thresholds were exceeded. SMS notifications were delivered promptly through the Twilio platform, enabling immediate awareness of critical events without requiring continuous manual observation. Experimental results indicate that the proposed framework achieved approximately 92% object detection accuracy while maintaining low processing latency suitable for real-time deployment. The integration of deep learning-based object detection with automated mobile notifications significantly enhances surveillance effectiveness by reducing response time and improving situational awareness. The proposed framework demonstrates excellent potential for deployment in intelligent transportation systems, smart cities, industrial safety monitoring, campus security, and public surveillance applications.

A. Dataset Description

Traffic recordings and image datasets that are accessible to the public were used to assess the system. The dataset roughly consists of:

- 100 video clips
- 500 image samples
- Resolution: 1280×720
- Frame rate: 30 FPS

The dataset covers various traffic conditions such as high-ways, intersections, and urban environments.

B. System Implementation

The Python programming language, YOLO-based deep learning models, and the Twilio communication service were used to create the suggested real-time object detection and alert system. Static photos and a number of traffic video clips were used to test the system. The assessment concentrated on the system's ability to identify cars, calculate their speed and distance, and send out alerts whenever potentially dangerous circumstances were identified. Figure 4 shows the graphical user interface of the developed system.

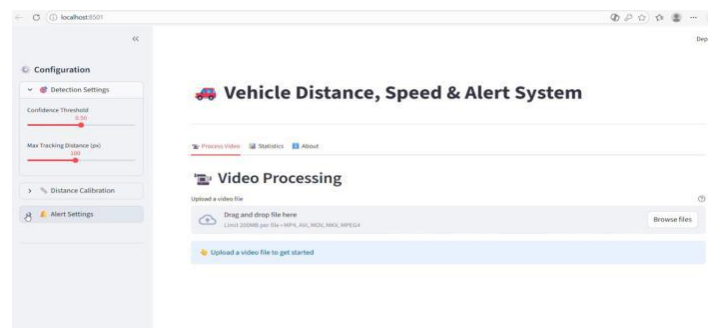


Figure 3: Graphical user interface of the developed vehicle distance, speed, and alert monitoring system

This interface allows users to run the video for real-time vehicle detection, upload video files, and modify parameters like the tracking distance and detection confidence level. Users may simply view the detection results thanks to the interface's straightforward and user-friendly design.

During video processing, the YOLOv8 model detects vehicles in each frame and draws bounding boxes around the identified objects. After detection, the system estimates the distance between vehicles by analyzing the size of the bounding boxes and their positions within the frame. The speed of each vehicle is then calculated by tracking its movement across consecutive frames. If the computed risk factor exceeds a predefined threshold, the system considers the situation potentially dangerous and automatically sends an alert message.

Figure 4 shows an example of the SMS alert notification generated by the system. The alert message includes important details such as the estimated distance and speed of the detected vehicles, helping users quickly understand the situation and take appropriate action.

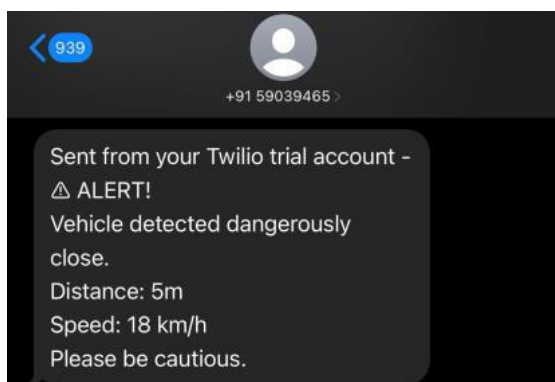


Figure 4: Example of SMS alert notification generated by the system using the Twilio communication platform

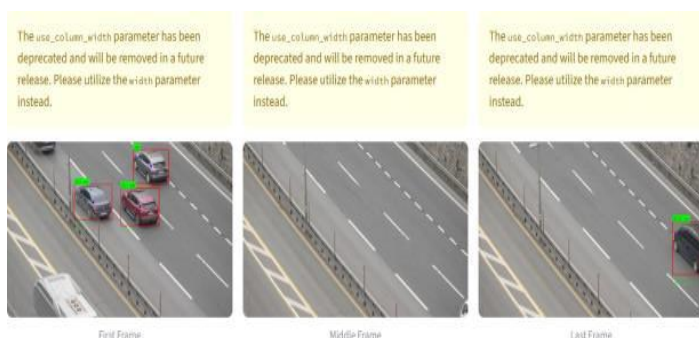


Figure 5: Vehicle detection results across three video frames (first, middle, and last)

The first frame shows multiple vehicles correctly detected with bounding boxes and labels. The middle frame contains no vehicles, demonstrating the absence of false positives. The last frame shows a single vehicle detected near the edge of the frame, indicating the model's ability to identify objects under varying conditions.

Figure 5 illustrates the vehicle detection performance across different frames of a traffic video. In the first frame, multiple vehicles are accurately identified with bounding boxes and confidence labels, demonstrating the model's ability to handle moderately dense traffic scenarios. The middle frame shows an empty road segment with no detections, confirming that the system effectively avoids false positives when no objects are present. In the last frame, a single vehicle is detected at the edge of the frame, highlighting the robustness of the model in handling partial visibility and varying object positions.

Furthermore, the results indicate that the proposed system maintains consistent detection accuracy across different traf- fic

conditions and frame variations. The model successfully adapts to changes in object count, spatial distribution, and scene complexity. The accurate localization of vehicles using bounding boxes also supports reliable distance and speed estimation in subsequent processing stages. These observations demonstrate that the system is capable of performing real-time detection while maintaining stability and precision, making it suitable for practical traffic monitoring and safety applications. Two demonstration videos are included to better illustrate the efficacy of the suggested strategy.

This video demonstrates how the system processes traffic video input in real time. Vehicles are detected using the YOLO model and highlighted with bounding boxes. During processing, the system continuously measures the distance and speed of the detected vehicles. If the computed risk factor exceeds the predefined threshold, the system automatically generates and sends an SMS alert to the user.

C. Evaluation Metrics

Standard object detection criteria, such as precision, recall, and mean Average Precision (mAP), were used to assess the system's performance.

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively. These metrics were used to evaluate the detection performance of the proposed system.

In addition, mean Average Precision (mAP) provides an overall measure of detection accuracy by considering both localization and classification performance. A higher mAP value indicates better detection capability.

The system successfully detected vehicles in both images and video streams while maintaining real-time performance. By combining YOLO-based detection with the Twilio messaging service, the system was able to generate and send alerts immediately when unsafe situations were detected.

Furthermore, the system maintained stable performance under different traffic conditions. The real-time processing ensures timely alert generation, demonstrating a good balance between accuracy and efficiency.

Table I: Performance Evaluation of Yolo Models Used in the Proposed System

Model	Precision	Recall	mAP	FPS
YOLOv3	0.88	0.86	0.87	30
YOLOv8	0.93	0.91	0.92	45

The experimental findings show that the suggested system provides excellent real-time performance and detection accuracy. When compared to traditional methods, the combination of distance and speed-based risk assessment greatly increases the accuracy of identifying dangerous circumstances.

In addition, the comparison results clearly indicate that YOLOv8 outperforms YOLOv3 in terms of precision, recall, and mAP, while also achieving a higher frame processing rate. This improvement highlights the effectiveness of using advanced deep learning models for real-time applications. The higher FPS value ensures smoother video processing and faster response times, which are critical for timely alert generation. Furthermore, the system maintains consistent performance across different traffic scenarios, demonstrating its reliability and practical applicability in real-world environments.

VI. FUTURE WORK

Although the proposed system demonstrates promising performance in real-time object detection and alert generation, there is still scope for further improvement. Detection accuracy can be enhanced by training the model on larger and more diverse datasets, allowing it to perform more reliably under different environmental conditions such as varying lighting, weather, and traffic density. In addition, the adoption of more advanced deep learning architectures, such as YOLOv9 or transformer-based detection models, can further improve detection precision, robustness, and processing efficiency. Although the proposed surveillance framework demonstrates effective real-time object detection and automated alert generation, several enhancements can further improve its performance and practical applicability. Future work may incorporate multi-object tracking algorithms such as DeepSORT or ByteTrack to provide more accurate object identity preservation and long-term movement analysis. The integration of transformer-based vision models and next-generation YOLO architectures may further improve detection accuracy for small, partially occluded, or densely packed objects.

The system may also be extended to detect abnormal human behavior, accidents, fire, smoke, weapon detection, and

unauthorized intrusions using action recognition and video understanding techniques. Integration with Internet of Things (IoT) devices, cloud computing platforms, and edge AI hardware would enable distributed surveillance with reduced latency and improved scalability. Furthermore, implementing GPS-based location tracking, real-time video streaming, multilingual voice alerts, mobile applications, and law enforcement integration would strengthen emergency response capabilities. Future research may also investigate privacy-preserving AI techniques, federated learning, and explainable artificial intelligence (XAI) to enhance user trust, data security, and regulatory compliance.

Future work may also focus on integrating additional sensing technologies, such as radar or LiDAR, to achieve more accurate and reliable distance estimation. This would help overcome the limitations of monocular vision-based approaches. Furthermore, the system can be extended for large-scale real-world applications, including smart city monitoring, advanced traffic management systems, and intelligent surveillance platforms. Such enhancements would increase the system's scalability, reliability, and practical usability in complex and dynamic environments.

VII. CONCLUSION

This study presents a real-time object detection and SMS alert system using the YOLO deep learning model and Twilio messaging service. The system detects vehicles and other objects from images and video streams, and identifies potentially dangerous situations such as close vehicle proximity. When such conditions occur, an alert message containing estimated distance and speed is generated and sent instantly to the user via SMS, enabling quick awareness and response. This integration ensures an efficient and automated monitoring process. This research presented an Intelligent Surveillance Using YOLO-Based Object Detection and Automated Mobile Alert System that integrates advanced deep learning techniques with cloud-based communication services to improve public safety and intelligent monitoring. The proposed framework employs YOLOv3 for accurate image-based object detection and YOLOv8 for high-speed real-time video analysis, enabling continuous detection of pedestrians, vehicles, and other objects under diverse environmental conditions. By incorporating distance estimation, speed analysis, intelligent risk assessment, and automated SMS notifications through the Twilio platform, the system provides immediate alerts whenever potentially dangerous situations are detected.

Experimental evaluation demonstrated that the proposed framework achieves approximately 92% object detection accuracy while maintaining real-time processing performance and reliable mobile alert delivery. The integration of computer vision, deep learning, and automated communication significantly reduces monitoring effort, shortens emergency response time, and enhances situational awareness. The proposed intelligent surveillance system offers a scalable, efficient, and cost-effective solution for traffic management, industrial safety, smart city infrastructure, campus security, and public surveillance. With future enhancements in AI-based behavior analysis, edge computing, and IoT integration, the framework has strong potential to support next-generation intelligent surveillance and emergency response systems. The results show that combining computer vision with cloud-based communication enables effective real-time monitoring and alert generation. The system achieves reliable detection performance with low latency and consistent accuracy under different traffic conditions. Its lightweight and cost-effective design makes it suitable for practical applications such as traffic monitoring and intelligent surveillance. Future work may focus on improving accuracy, integrating additional sensors, and extending the system for large-scale smart city environments.

REFERENCES

- [1] Sarkar, M., Sulfi, S., Rifaz, M. A. S., Aswin, A., & Shefi, T. M. (2024). Accident Detection and Alert System Using YOLOv8 and Twilio. *International Journal of Creative Research Thoughts (IJCRT)*, 12(4), c231–c234.
- [2] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779–788.
- [3] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6517–6525.
- [4] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*.
- [5] Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv preprint arXiv:2004.10934*.
- [6] Jocher, G., Chaurasia, A., & Qiu, J. (2023). YOLOv5 by Ultralytics. *GitHub Repository*. <https://github.com/ultralytics/yolov5>.
- [7] Jocher, G., et al. (2023). YOLOv8: Next Generation Object Detection Model. *Ultralytics Documentation*.
- [8] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149.
- [9] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. *European Conference on Computer Vision (ECCV)*, 21–37.
- [10] Rehman, S. U., Khan, S. A., Arif, A., & Khan, U. S. (2021). IoT-Based Accident Detection and Emergency Alert System for Motorbikes. *International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Islamabad, Pakistan.
- [12] Wani, L. K., Momin, M. M., Bhosale, S., Yadav, A., & Nili, M. (2022). Vehicle Crash Detection Using YOLO Algorithm. *International Journal of Computer Science and Mobile Computing*, 11(5), 24–30.
- [13] Darokar, A., Talware, A., Jadhav, A., Guntiwari, D., & Dandhare, S. (2023). Accident Detection System Using YOLOv8 and CNN Algorithm. *International Journal of Innovative Research in Engineering and Technology*, 12(4), 234–239.
- [14] Bhingare, D., Gadapele, M., Gautam, J., Kakade, T., & Thakre, A. (2022). Emergency Alert System Using IoT. *International Journal of Advanced Research in Computer and Communication Engineering*, 11(3), 45–49.
- [15] Lou, H., Duan, X., Guo, J., Liu, H., Gu, J., Bi, L., & Chen, H. (2023).
- [16] DC-YOLOv8: Small Size Object Detection Algorithm Based on Camera Sensor. *Journal of Intelligent Systems*, 32(4), 1–10.
- [17] Huang, X., Qiao, H., Li, H., & Jiang, Z. (2022). Bioinspired Approach- Sensitive Neural Network for Collision Detection in Dynamic Environments. *Applied Soft Computing*, 122.
- [18] Li, T., & Stern, R. (2021). Classification of Adaptive Cruise Control Vehicle Type Based on Car-Following Trajectories. *IEEE Intelligent Transportation Systems Conference (ITSC)*, 1547–1552.
- [19] Chen, Y., Zhang, X., & Liu, Y. (2020). Deep Learning-Based Intelligent Surveillance System for Traffic Monitoring. *IEEE Access*, 8, 182123–182134.
- [20] Kumar, S., Singh, R., & Sharma, P. (2021). Real-Time Traffic Monitoring System Using Deep Learning

- Techniques. International Journal of Computer Applications, 174(15), 22–28.
- [21] Zhang, Y., Wang, H., & Li, J. (2022). Intelligent Surveillance System for Object Detection and Classification Using Deep Learning. Journal of Artificial Intelligence Research, 73, 455–472.
- [22] Singh, A., Patel, R., & Verma, S. (2021). Smart Monitoring System Using Computer Vision and IoT Technologies. International Journal of Emerging Technologies in Engineering Research, 9(6), 112–118.
- [23] T. K. Dutta and S. Kar, “On Regular Ternary Semiring,” in Advances in Algebra, Proceedings of the ICM Satellite Conference in Algebra and Related Topics, World Scientific, 2003, pp. 343–355.
- [24] T. K. Dutta and M. L. Das, “On Strongly Prime Semiring,” Bull. Malays. Math. Sci. Soc., vol. 30, no. 2, pp. 135–141, 2007.
- [25] D. Handelman and J. Lawrence, “Strongly Prime Rings,” Trans. Amer. Math. Soc., vol. 211, pp. 209–223, 1975.
- [26] W. G. Lister, “Ternary Rings,” Trans. Amer. Math. Soc., vol. 154, pp. 37–55, 1971.
- [27] M. Sajani Lavanya, D. Madhusudhana Rao, and V. Syam Julius Rajen- dra, “On Lateral Ternary Γ -Ideals of Ternary Γ -Semirings,” American International Journal of Research in Science, Technology, Engineering & Mathematics, vol. 12, no. 1, pp. 11–14, 2015.
- [28] M. Sajani Lavanya, D. Madhusudhana Rao, and V. Syam Julius Rajen- dra, “On Quasi-Ternary Ideals and Bi-Ternary Ideals in Ternary Semirings,” International Journal of Mathematics and Statistics Invention, vol. 3, no. 6, pp. 5–14, 2015.
- [29] M. Sajani Lavanya, D. Madhusudhana Rao, and V. Syam Julius Rajen- dra, “Prime Bi-Ternary Ideals in Ternary Semirings,” British Journal of Research, vol. 2, no. 6, pp. 156–166, 2015.
- [30] M. Sajani Lavanya, D. Madhusudhana Rao, and V. Syam Julius Rajen- dra, “A Study on the Jacobson Radical of a Ternary Semiring,” International Journal of Mathematics and Computer Applications Research, vol. 6, no. 1, pp. 17–30, 2016.
- [31] G. Srinivasa Rao, D. Madhusudhana Rao, and P. Siva Prasad, “Ideals in Quotient Ternary Semiring,” International Journal of Advances in Management, Technology and Engineering Sciences, vol. 7, no. 12, pp. 126–134, 2017.
- [32] G. Srinivasa Rao, P. Siva Prasad, M. Vasanth, and D. Madhusudhana Rao, “On Strongly Duo and Duo Left Γ -TS-Acts over ternary semi- groups,” International Journal of Pure and Applied Mathematics, vol. 113, no. 6, pp. 65–73, 2017.
- [33] Ch. Manikya Rao, P. Siva Prasad, D. Madhusudhana Rao, and G. Srinivasa Rao, “Maximal Ideal of Compact Connected Topological Ternary Semigroups,” in International Conference on Mathematics, Kerala, 2015.
- [34] P. Siva Prasad, D. Madhusudhana Rao, and G. Srinivasa Rao, “A Study on Structure of PO-Ternary Semirings,” Journal of Advances in Mathematics, vol. 10, no. 8, pp. 3717–3724.
- [35] P. Siva Prasad, D. Madhusudhana Rao, M. Vasanth, and B. Srinivasa Kumar, “On Γ -TS-Acts over Ternary Γ -Semigroups,” International Journal of Engineering & Technology, vol. 7, no. 4.10, pp. 812–815, 2018.
- [36] P. Siva Prasad, K. Revathi, P. Sundarayya, and D. Madhusudhana Rao, “Compositions of Fuzzy T-Ideals in Ternary Semiring,” International Journal of Advances in Management, Technology and Engineering Sciences, vol. 7, no. 12, pp. 135–145, 2017.
- [37] P. Siva Prasad, C. Sreemannarayana, D. Madhusudhana Rao, T. Nageswara Rao, and K. Anuradha, “On Left Ternary Semigroups-I,” International Journal of Recent Technology and Engineering, vol. 7, ICETESM, pp. 165–167, 2019.
- [38] P. Siva Prasad, C. Sreemannarayana, D. Madhusudhana Rao, T. Nageswara Rao, and M. Sajani Lavanya, “On Left Ternary Semigroups- I,” International Journal of Recent Technology and Engineering, vol. 7, ICETESM, pp. 168–170, 2019.

Citation of this Article:

Carlo Miguel Garcia, & Kevin Angelo Torres. (2026). Intelligent Surveillance Using YOLO-Based Object Detection and Automated Mobile Alert System. *Current Journal of Engineering and Science Research*. 3(6), 51-65. Article DOI: <https://doi.org/10.47001/CJESR/2026.306006>

***** End of the Article *****