International Journal of Research and Review
[Indexed, Peer-Reviewed Journal]
Volume 13; Issue: 4; April 2026
Website: www.ijrrjournal.com
Research Paper *E-ISSN: 2349-9788; P-ISSN: 2454-2237*

Surveillance System Using Deep Learning with Automatic Report Generation

Putti Venkata Siva Teja[1], Sirla Sowmya[1], Shaik Naseema Banu[2], Tungala Nandu Sree[3], Sangepu Venkat Sai[4], Shaik Mohammed Irfan[5], Kondapalli Devendranadh[6]

[1,2,3,4,5,6]Department of Information Technology,
Dhanekula Institute of Engineering & Technology, Vijayawada, Andhra Pradesh, India.

Corresponding Author: Putti Venkata Siva Teja

DOI: https://doi.org/10.52403/ijrr.20260411

ABSTRACT

With the increasing deployment of surveillance cameras across cities and public spaces, the massive volume of generated video data has made manual monitoring time-consuming, inefficient, and prone to human error. To address these challenges, this paper presents an intelligent surveillance system that leverages deep learning techniques to automatically analyze video streams and generate structured textual reports. The proposed framework integrates multiple computer vision components, including YOLOv8 for detecting dangerous objects such as knives and blood stains, along with Convolutional Neural Networks (CNNs) for real-time face recognition and activity suspicious activity detection. A webcam is used to capture live video input, enabling continuous monitoring and visualization of the environment. OpenCV (Open-Source Computer Vision) is employed to efficiently process video frames and display detection results in real time. A key innovation of this system is the Automatic Report Generation module, which utilizes natural language generation to convert detected events into summaries containing precise details such as event snaps, location, and timestamps. Furthermore, the system incorporates a proactive security approach by integrating an ESP32-controlled hardware mechanism for remote access control and mobile alerts via a Flutter application. Experimental evaluations demonstrate high performance, with detection accuracies reaching 95.1% for unknown persons and 94.2% for weapons, alongside a rapid report generation response time of 25 ms. This end-to-end solution minimizes the need for continuous human supervision while significantly improving the reliability and responsiveness of surveillance operations in smart cities and industrial environments.

Keywords: Deep Learning, Intelligent Surveillance System, OpenCV, YOLOv8, Convolutional Neural Network (CNN), Suspicious activity Detection, Automatic Report Generation, ESP32.

INTRODUCTION

Surveillance systems are very important in keeping our world a safer place. This is why there is a very big need for systems that can be used independently. This is where our research comes in, where we developed a real-time image detection tool that uses a very smart artificial intelligence model known as YOLOv8, a very fast and accurate model of artificial intelligence [1]. These surveillance systems are used in places such as airports, train stations, shopping centres, schools, and even homes. Their main

purpose is to keep watch, prevent crimes, and act as evidence in case of a crime [2], [3]. With a rapid increase in the number of CCTV cameras, a significant amount of video data is being generated daily. Analysing this data manually is becoming increasingly cumbersome. Conventional surveillance systems require a lot of human involvement, and there is a high possibility of missing critical events due to distractions or a lack of attention span. This may delay action in critical situations, affecting overall security. Recent advantage in deep learning techniques and computer vision have enabled the development of intelligent surveillance systems that can automatically analyse the video data. The addition of OpenCV has further improved its efficiency, enabling it to be used in complex environments. These systems can detect weapons, recognize faces, detect unusual activities, and analyse crowd behaviour. Another vital aspect of smart surveillance systems is the generation of reports. Rather than spending more time reviewing surveillance footage, the system can generate reports indicating what happened, when, and where [6], [7]. Intelligent surveillance systems can also be effective in providing security through access control systems like servo motor door lock systems, as well as through mobile alerts [3], [6]. For example, in the event of detecting an unknown person in the area, the system can send alerts in real-time, enabling security personnel to act immediately [2], [3]. The proposed system also improves surveillance by detecting dangerous objects like knives and blood stains, crowd formation, detecting authorized or unauthorized persons through face recognition [2], [5]. Multitask CNN improves Multiview face detection accuracy; however, the recall depends on the initial detection window [8], [9]. In conclusion, intelligent surveillance systems based on deep learning reduce human effort significantly, improving accuracy in surveillance [3], [4]. The proposed intelligent surveillance system can be considered complete with the addition of all these features. This work was based on developing a unified intelligent surveillance system in real-time, integrating face detection, face recognition, alerts, and report generation [1]– [9].

PROPOSED SYSTEM

The proposed intelligent surveillance framework is designed as a modular, layered architecture that facilitates seamless video acquisition, real-time deep learning analysis, and automated reporting. The system initiates with a Video Acquisition Module that captures live streams and transitions them into a Preprocessing Module, where OpenCV is utilized for noise reduction and normalization to ensure high-quality input for neural network inference. At the core of the architecture, the Deep Learning Analysis Module employs YOLOv8 for the high-precision detection of dangerous objects, such as knives and blood stains, while parallel units perform facial recognition and crowd density monitoring. A defining feature of this system is the integration of an Automatic Report Generation module, which applies natural language generation techniques to convert structured event data—including timestamps, location, and confidence scores—into human-readable summaries for immediate documentation. Furthermore, the system bridges the gap between software and physical security through an IoT-based hardware layer; an ESP32 microcontroller manages a servo-controlled locking mechanism, allowing security personnel to remotely grant or deny access via a Flutter mobile application based on real-time threat assessments.

SYSTEM ARCHITECTURE

The architecture of the proposed Surveillance System Using Deep Learning with Automatic Report Generation is designed as a modular and layered framework that enables efficient video acquisition, intelligent analysis, and automated reporting as shown in figure 1. Each module performs a specific function

while interacting seamlessly with other components to ensure smooth end-to-end operation. At the lowest level, the system begins with the Video Acquisition Module, which captures live video streams from surveillance cameras. These video streams are transmitted to the processing unit, where they are converted into frames and forwarded for further analysis. This module supports both real-time camera feeds and stored video inputs, providing flexibility for different deployment scenarios. The captured frames are passed to the Preprocessing Module, where noise reduction, resizing, normalization, and frame enhancement operations are performed. Preprocessing improves image quality and ensures that the input data is in a suitable format for deep learning models. This step is essential for achieving consistent and reliable detection results. Next, the Deep Learning Analysis Module forms the core of the system. This module consists of three major sub- components: Object Detection Unit, which identifies and localizes objects such as humans, and other relevant entities in each frame. Face Recognition Unit, which extracts facial features and compares them with stored templates to identify known individuals. Activity Recognition Unit, which analyzes temporal patterns across frames to recognize human actions and behaviors. The outputs of these units are combined and forwarded to the Event Interpretation Module, which correlates detected objects, recognized faces, and identified activities to determine meaningful events. For example, the presence of an unknown person in a restricted area can be interpreted as a suspicious event. Once events are identified, the Alert and buzzer activation and Generation Module trigger mobile notifications in case of abnormal or critical situations. Alerts notifying to security personnel through predefined channels. Finally, the Automatic Report Generation Module converts event information into structured textual reports. These reports include details such as event snap, time stamp, location, detected entities, and confidence scores. The generated reports are stored in a database and can be accessed later for auditing, investigation, or analysis purposes. Overall, the modular architecture ensures scalability, flexibility, and ease of integration. New models or functionalities can be added without affecting the entire system, making the architecture suitable for real-world large- scale surveillance applications.

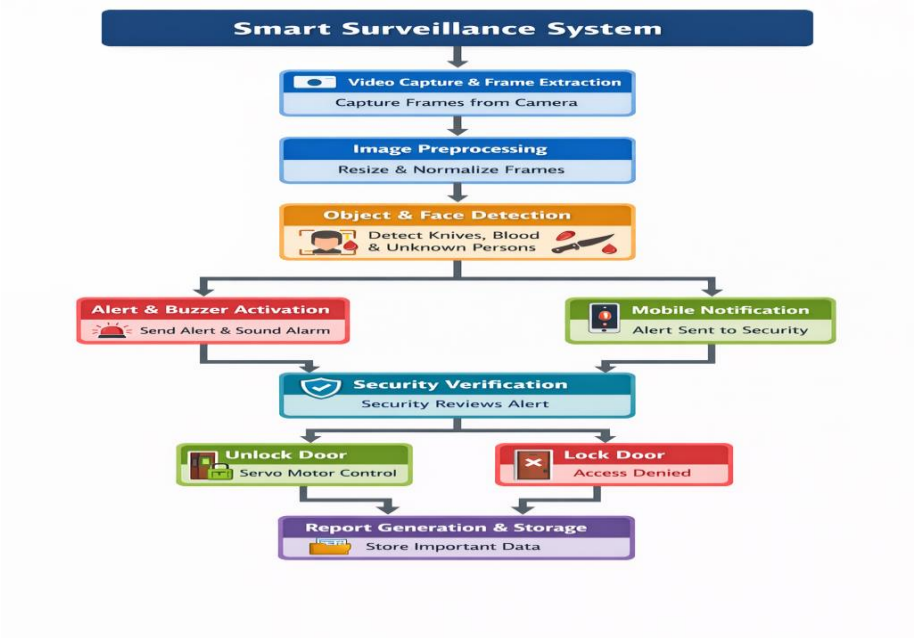

Figure 1: System Architecture

DATA SET

The performance of deep learning-based surveillance systems depends heavily on the quality and diversity of training data. The proposed system utilizes multiple publicly available datasets along with custom-collected video samples to ensure robust model training and evaluation. The object detection model is trained using datasets containing labeled images of common objects such as people, and everyday items. As shown in Table 1, These datasets include thousands of images captured under different lighting conditions, viewpoints, and backgrounds. For face recognition, datasets containing large numbers of facial images from multiple individuals are used. These datasets include variations in pose, expression, age, and illumination, enabling the model to learn discriminative facial features. Activity recognition models are trained on video datasets that contain annotated human actions. Each video clip is labeled with a specific activity class, such as walking, running, sitting, or fighting. This allows the model to learn temporal patterns associated with different behaviors. In addition to public datasets, a small custom dataset is created by capturing videos in controlled environments. This dataset is used to fine-tune the models and evaluate system performance under real-world conditions. All datasets are divided into training, validation, and testing subsets. Data augmentation techniques such as rotation, flipping, scaling, and brightness adjustment are applied to increase diversity and prevent overfitting.

Table 1: Datasets Used

Dataset	Task	Data Type	Size
COCO	Object Detection	Images	330K+
LFW	Face Recognition	Images	13K+
VGGFace2	Face Recognition	Images	3M+
UCF101	Activity Recognition	Videos	13K+
HMDB51	Activity Recognition	Videos	7K+
Custom Dataset	System Testing	Videos	500+ clips

RESULTS & DISCUSSION

The proposed intelligent surveillance system was evaluated using multiple test videos and real-time camera feeds to assess its effectiveness in detecting weapons, blood stains, unknown persons, and crowd situations, as well as its ability to generate automatic reports. The evaluation focused on detection accuracy, response time, and overall system reliability. As shown in Table 2 & Table 3, The object detection models successfully identified knives and blood stains in different lighting conditions and backgrounds. Face recognition experiments demonstrated reliable identification of known individuals and accurate detection of unknown persons. Crowd detection results showed that the system can effectively determine overcrowded scenes by counting detected faces and comparing them with predefined thresholds. The automatic report generation module consistently produced structured textual summaries for each detected event. These reports included event type, time stamp, camera ID, and confidence score, which improved situational awareness and reduced the need for manual documentation.

Table 2: Detection Accuracy of Proposed System

Detection Task	Accuracy (%)
Knife Detection	94.2
Blood Stain Detection	92.8
Unknown Person Detection	95.1
Crowd Detection	93.5

Table 3: Response Time

Module	Response Time (ms)
Knife Detection	45
Blood Stain Detection	48
Face Recognition	60
Crowd Detection	40
Report Generation	25

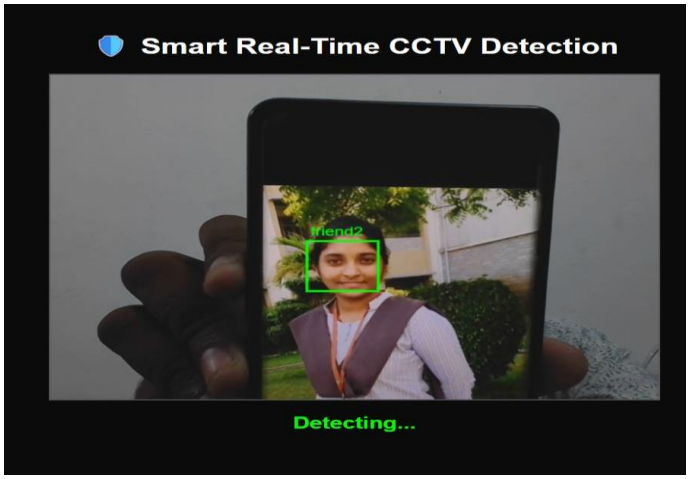

Figure 2: Face Detection as known

As shown in Figure 2 The face detection and recognition component of the system functions as a critical security layer by processing live video streams to identify individuals in real time. Upon capturing the video feed, the system utilizes deep learning algorithms to isolate human faces and encapsulate them within precise bounding boxes. Once a face is localized, the system extracts unique facial embeddings and performs a high-speed comparison against a pre-registered database. This matching process ensures accurate identification, allowing the system to distinguish between authorized personnel and unknown visitors, which subsequently informs the automated access control and reporting modules.

Figure 3: Face Detection unknown

In the event that the system captures an individual not present in the authorized database, the facial recognition unit categorizes the subject as "Unknown." As shown in figure 3. The system immediately isolates the unidentified face within a highlighted bounding box and captures a high-resolution image for storage and future forensic reference. Simultaneously, the framework triggers an automated security protocol, generating a real-time alert—such as a localized buzzer notification or a mobile push alert—to ensure immediate intervention by security personnel.

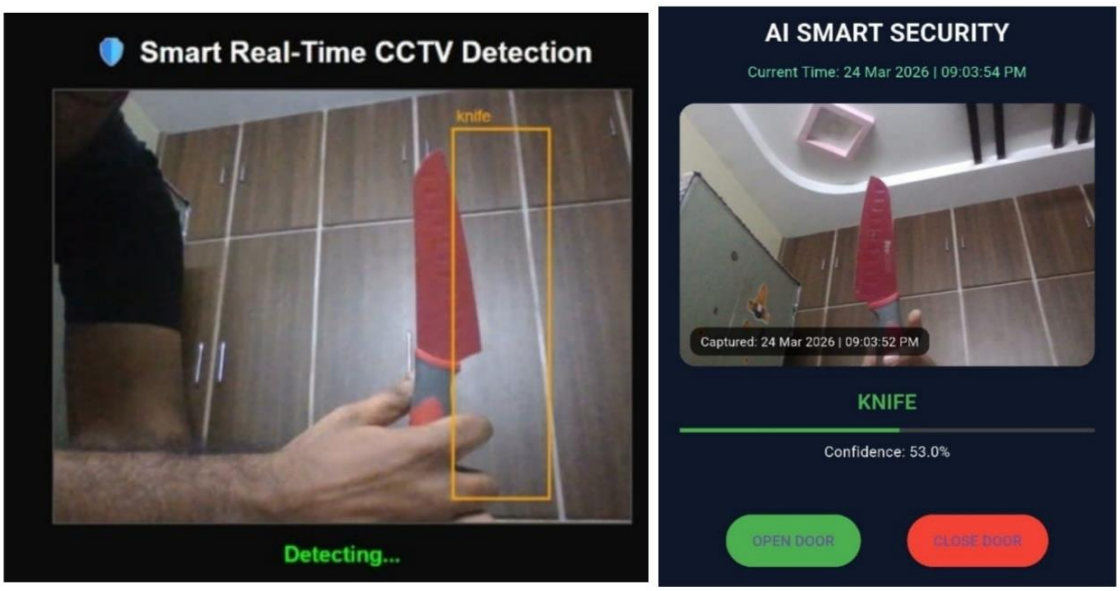

Figure 4: Knife Detection

The weapon detection module significantly enhances the system's proactive security capabilities by continuously monitoring live video feeds for high-risk objects. As shown in figure 4 Specifically, the system is engineered to identify knives and other sharp implements in real time as they appear within the camera's field of view. Upon detection, a precise bounding box is rendered around the object, labeled with a high-confidence classification to distinguish the threat from non-hazardous items. By accurately identifying these sharp objects as potential threats, the system provides a reliable automated surveillance layer that ensures critical incidents are flagged immediately for human intervention or automated lockdown protocols.

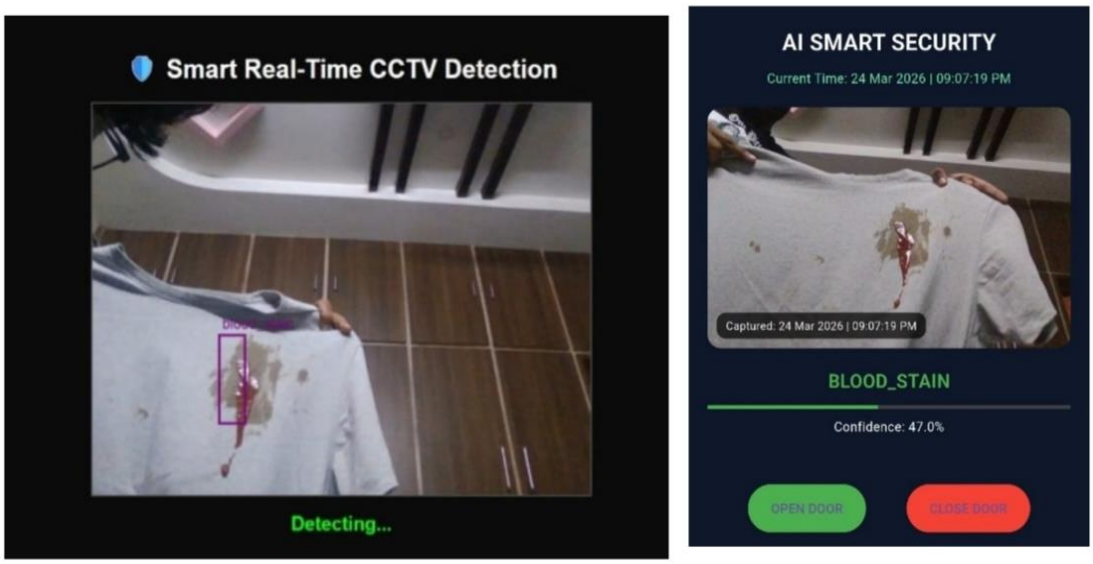

Figure 5: Blood Stain Detection

Beyond weapon identification, the system incorporates a specialized forensic detection layer focused on identifying biological markers of distress, specifically blood stains. By analyzing the live video feed from the connected webcam, the deep learning model can distinguish blood-red pixel patterns and textures from common background objects in real time. Upon detection, the system overlays a localized

bounding box and applies a corresponding label to the interface, ensuring that the event is captured without latency. As shown in figure 5, This capability is particularly vital for emergency response scenarios, as it allows the system to automatically flag violent encounters or accidents that require immediate medical or security intervention, thereby proving the system's comprehensive utility in high-stakes safety and security monitoring.

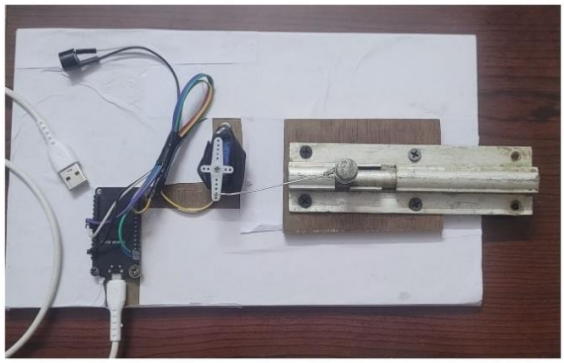

Figure 6: Smart door locking System (door open)

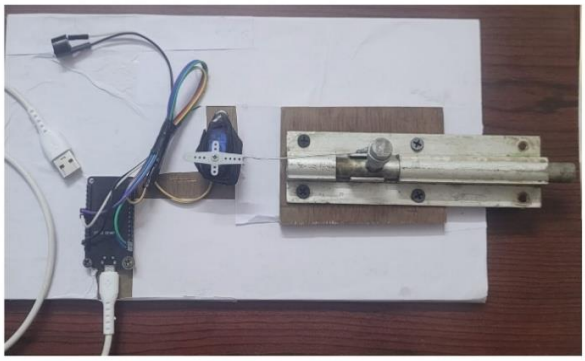

Figure 7: Smart locking System (door close)

The hardware integration of the system provides a physical response layer to the digital detection outputs through an automated access control mechanism. As illustrated in Figure 6 and Figure 7, a Servo Motor is employed to manage the state of the door lock, transitioning between open and closed positions based on system logic. This motor is interfaced with an ESP32 Development Board, which acts as the central processing unit for hardware signals, interconnected via high-quality jumper wires to the motor and a synchronized buzzer. When the facial recognition module identifies an authorized individual, the ESP32 triggers the servo motor to rotate, successfully unlocking the door as shown in Figure 6. Conversely, as depicted in Figure 7, if an unknown or suspicious person is detected, the system maintains a secure "door closed" state; simultaneously, the ESP32 activates the buzzer to provide an audible security alert. The experimental results validate this integration, confirming that the system achieves high detection accuracy across all parameters and summary of events that can be used for efficient decision-making whether the person is allow or not allow. The deep learning models for knife and blood stain detection remain reliable even in complex visual scenes, while the face recognition unit demonstrates a robust capability in distinguishing known users from intruders. With an average response time of 25 ms for report generation and minimal latency for hardware triggers, the system ensures that physical interventions and structured event documentation occur in real time, delivering an accurate, fast, and reliable security solution.

CONCLUSION

The present paper proposes an intelligent surveillance system that incorporates deep learning techniques and report generation to improve security monitoring. It can detect critical events like the detection of knives, blood stains, unknown persons, and crowd conditions in real time. By integrating object detection, face recognition, and basic activity detection with report generation, this system can improve the efficiency of monitoring activities. The experimental outcome of this system indicates that it can be implemented in real-time scenarios to achieve high detection accuracy and response times. By integrating report generation in this system, it has added more value to the intelligence of this system by providing a summary of events that can be used for efficient decision-making. Based

on the experimental outcome and report generation integration, this system can be considered a more efficient solution to meet the requirements of modern surveillance systems. In future, this system can be improved by integrating advanced activity detection techniques and more sophisticated report generation techniques to improve its intelligence.

Declaration by Authors

Acknowledgement: None
Source of Funding: None
Conflict of Interest: No conflicts of interest declared.

REFERENCES

1. Pavani Chitrapu, Mahesh Kumar Morampudi and Hemantha Kumar Kalluri, "Robust Face Recognition Using Deep Learning and Ensemble Classification," in IEEE Access, vol. 13, pp. 99957-99969, 2025, doi: 10.1109/ACCESS.2025.3575192.
2. Amulya Reddy Maligireddy, Manohar Reddy Uppula, Nidhi Rastogi, and Yaswanth Reddy Parla "Gun Detection Using Combined Human Pose and Weapon Appearance," arXiv preprint arXiv:2503.12215, 2025. doi:10.48550/arXiv.2503.12215.
3. Nashwan Adnan Othman, Mustafa Zuhaer Nayef Al-Dabagh and Ilhan AYDIN, "A New Embedded Surveillance System for Reducing COVID-19 Outbreak in Elderly Based on Deep Learning and IoT," 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI), 2020, pp. 1-6, doi: 10.1109/ICDABI51230.2020.9325651.
4. Muhammad Javed Iqbal, Muhammad Munwar Iqbal, Iftikhar Ahmad, Madini O. Alassafi, Ahmed S. Alfakeeh, and Ahmed Alhomoud, "Real-Time Surveillance Using Deep Learning," Security and Communication Networks, 2021, Art. no. 6184756, doi: 10.1155/2021/6184756.
5. Jyothi Kukad, Swapnil Soner, and Sagar Pandya, "Autonomous anomaly detection system for crime monitoring and alert generation," Journal of Automation, Mobile Robotics, and Intelligent Systems, 2022, doi: 10.14313/JAMRIS/1-2022/7.
6. Ruchi Rani, Kiran Napte, Sumit Kumar, Sanjeev Kumar Pippal, and Megha Dalsaniya, "Face recognition system for criminal identification in CCTV footage using Keras and OpenCV," Ingénierie des Systèmes d'Information, vol. 30, no. 3, p. 647, 2025, doi: 10.18280/isi.300309.
7. Mohammad Zahrawi and Khaled Shaalan, "Improving video surveillance systems in banks using deep learning techniques," Scientific Reports, vol. 13, Art. no. 7911, 2023, doi: 10.1038/s41598-023-35190-9.
8. Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499-1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.
9. Celia Cabello-Collado, Javier Rodriguez-Juan, David Ortiz-Perez, Jose Garcia-Rodriguez, David Tomás, and Maria Flores Vizcaya-Moreno, "Automated generation of clinical reports using sensing technologies with deep learning techniques," Sensors, vol. 24, no. 9, p. 2751, Apr. 2024, doi:10.3390/s24092751.

How to cite this article: Putti Venkata Siva Teja, Sirla Sowmya, Shaik Naseema Banu, Tungala Nandu Sree, Sangepu Venkat Sai, Shaik Mohammed Irfan et al. Surveillance system using deep learning with automatic report generation. *International Journal of Research and Review*. 2026; 13(4): 101-108. DOI: *https://doi.org/10.52403/ijrr.20260411*

******