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 6, Issue 14, April 2026
Impact Factor: 8.2

AI-Based Smart Surveillance System with Multi-Tenant Architecture, Intruder Detection, and Real-Time Alerting

Samarth S. Kulkarni[1], Varad V. Lokare[1], Prasad S. Gaikwad[1], Kartik S. Shinde[1], Dhruva V. Gajbhiye[1], Mr. A. D. Bagale[2]

[1]Department of Computer Engineering, SVERI's College of Engineering, Pandharpur. India
[2] Assistant Professor, SVERI's College of Engineering, Pandharpur. India

Abstract: *This paper presents an AI-based smart surveillance system that integrates object detection, facial recognition, and scalable backend architecture to improve security systems. The system utilizes YOLOv8 for real-time human detection and DeepFace for identity recognition. A key feature is the implementation of a multi-tenant architecture that allows different institutions to use different databases such as MySQL, PostgreSQL, or cloud-based systems through a unified platform. When an unknown individual is detected, the system captures images, generates sketch-like outputs, and sends real-time email alerts. Additionally, FastAPI is used to provide API endpoints for monitoring and integration. The system supports CCTV and IP camera inputs, making it practical for real-world deployment. Experimental results show that the system performs efficiently in real-time environments with high accuracy and scalability.*

Keywords: Smart Surveillance System, Computer Vision, YOLOv8, Facial Recognition, DeepFace, Intruder Detection, Multi-Tenant Architecture, Real-Time Monitoring, Object Detection, Image Processing, FastAPI, Security Systems

I. INTRODUCTION

In recent years, the rapid advancement of technology and the increasing need for public and private security have led to the widespread deployment of surveillance systems across various environments such as educational institutions, banks, airports, corporate offices, and smart cities. Traditional surveillance systems primarily rely on Closed-Circuit Television (CCTV) cameras, which continuously capture video footage for monitoring purposes. However, these systems depend heavily on human operators to observe and analyze video streams, making them inefficient, labor-intensive, and prone to human error. Continuous manual monitoring is not only exhausting but also increases the chances of missing critical events, especially in high-density or multi-camera environments. As a result, there is a growing demand for intelligent surveillance systems that can automatically detect, analyze, and respond to suspicious activities in real time.

Artificial Intelligence (AI), particularly in the field of computer vision, has emerged as a powerful solution to address these challenges. Computer vision enables machines to interpret and understand visual data from the world, making it possible to automate tasks such as object detection, facial recognition, and activity analysis. Among the various techniques available, deep learning-based models have shown remarkable performance in visual recognition tasks. Models such as You Only Look Once (YOLO) have revolutionized real-time object detection by offering high accuracy and speed, making them suitable for surveillance applications. Similarly, facial recognition systems powered by deep learning frameworks such as DeepFace and FaceNet have significantly improved the ability to identify individuals with high precision.

DOI: 10.48175/IJARSCT-34068

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 6, Issue 14, April 2026 Impact Factor: 8.2

Despite these advancements, many existing surveillance systems still suffer from several limitations. Most systems are designed for a single environment and lack scalability, making it difficult to deploy them across multiple organizations or institutions. Additionally, traditional systems often use fixed database configurations, which restrict flexibility and adaptability. Another major limitation is the absence of real-time alert mechanisms, which are crucial for immediate response in critical situations. Furthermore, many systems do not provide external access or integration capabilities, limiting their usability in modern interconnected environments.

To overcome these limitations, this paper proposes an AI-based smart surveillance system that integrates real-time object detection, facial recognition, and a scalable backend architecture. The system is designed to automatically detect human presence using YOLOv8, extract facial features, and recognize individuals using DeepFace. Based on the recognition results, individuals are classified as either known or unknown. Known individuals are verified against a pre-existing database, while unknown individuals are treated as potential intruders.

One of the key contributions of this work is the implementation of a multi-tenant database architecture. This approach allows multiple institutions or organizations to use the same surveillance system while maintaining separate and independent databases. For example, one institution can use MySQL, another can use PostgreSQL, and a third can use a cloud-based database, all within the same system framework. This flexibility significantly enhances the scalability and adaptability of the system, making it suitable for real-world deployment across diverse environments.

In addition to detection and recognition, the proposed system incorporates intelligent intruder handling mechanisms. When an unknown individual is detected, the system not only logs the event but also captures an image of the intruder and generates a sketch-like representation using image processing techniques. This feature is particularly useful in scenarios where the captured image quality is poor or unclear, as it provides an alternative visual representation that can aid in identification. Furthermore, the system includes a real-time alert mechanism that sends email notifications to administrators whenever an intruder is detected. This ensures immediate awareness and enables quick response to potential security threats.

Another important feature of the system is the integration of an API layer using FastAPI. This allows external systems, dashboards, or applications to access surveillance data in real time. Through these APIs, users can retrieve logs, monitor events, and analyze system performance. This capability enhances the overall usability and extensibility of the system, making it compatible with modern web-based monitoring solutions.

The system is designed to be compatible with multiple input sources, including webcams, CCTV cameras, and IP camera streams. This ensures that it can be easily integrated into existing surveillance infrastructures without requiring significant modifications. The modular design of the system further allows for easy upgrades and future enhancements, such as cloud deployment, mobile application integration, and advanced analytics.

In summary, the proposed AI-based smart surveillance system aims to address the limitations of traditional surveillance by providing an automated, scalable, and intelligent solution. By combining real-time detection, facial recognition, multi-tenant architecture, and alert mechanisms, the system enhances security, reduces human dependency, and improves response efficiency. The integration of modern technologies such as deep learning and API-based communication makes the system highly adaptable and suitable for deployment in a wide range of real-world scenarios.

II. LITERATURE REVIEW

The development of intelligent surveillance systems has gained significant attention in recent years due to advancements in computer vision and deep learning techniques. Object detection plays a crucial role in surveillance, and models such as You Only Look Once (YOLO) have become widely popular for real-time applications. Introduced by Redmon et al., YOLO revolutionized object detection by framing it as a single regression problem, enabling high-speed and accurate detection. Subsequent versions, including YOLOv4 and YOLOv8, have further improved detection accuracy and efficiency, making them suitable for real-time surveillance environments.

In addition to object detection, facial recognition is a key component of modern surveillance systems. Deep learning-based approaches such as FaceNet and DeepFace have demonstrated remarkable performance in identifying

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 6, Issue 14, April 2026 Impact Factor: 8.2

individuals. FaceNet, proposed by Schroff et al., introduced the concept of learning a unified embedding space for faces, allowing efficient comparison using distance metrics. Similarly, DeepFace utilizes deep neural networks to extract facial features and achieve near-human-level accuracy in face verification tasks. These advancements have made it possible to reliably recognize individuals in dynamic environments.

Several existing surveillance systems combine object detection and facial recognition to automate monitoring. However, many of these systems are limited in terms of scalability and flexibility. Most traditional implementations are designed for a single organization and rely on fixed database structures, which restrict their ability to adapt to different deployment scenarios. Furthermore, many systems lack real-time alert mechanisms, which are essential for immediate response to security threats. The absence of integration capabilities, such as APIs, also limits their usability in modern interconnected systems.

Recent research has explored the use of cloud-based and distributed architectures to enhance scalability in surveillance systems. However, these approaches often introduce additional complexity and cost. The concept of multi-tenant architecture, commonly used in software systems, has not been widely applied in surveillance solutions. Multi-tenant systems allow multiple users or organizations to share a single application while maintaining separate data environments, thereby improving resource utilization and flexibility.

The proposed system builds upon these existing technologies by integrating real-time object detection, facial recognition, and a multi-tenant database architecture into a unified framework. It also incorporates additional features such as image capture, sketch generation, email alerts, and API-based data access. By addressing the limitations of existing systems, this work contributes to the development of a more scalable, flexible, and intelligent surveillance solution suitable for real-world applications.

III. METHODOLOGY

The proposed smart surveillance system is designed as a modular and scalable pipeline that integrates real-time object detection, facial recognition, event logging, and alert mechanisms. The methodology focuses on processing live video streams, identifying human presence, recognizing individuals, and taking appropriate actions based on classification results. The system operates in a sequential yet optimized manner to ensure real-time performance and accuracy.

The first stage of the methodology involves input acquisition, where video data is captured from various sources such as webcams, CCTV cameras, or IP-based camera streams. The flexibility in input sources allows the system to be deployed in multiple real-world environments without requiring major infrastructure changes. Each frame from the video stream is processed individually to maintain continuous monitoring.

Once the input frame is obtained, it is passed to the human detection module. This module utilizes the YOLOv8 (You Only Look Once version 8) deep learning model, which is known for its high speed and accuracy in object detection tasks. YOLOv8 processes the entire image in a single forward pass and identifies objects of interest, particularly humans in this case. The model outputs bounding boxes around detected persons along with confidence scores. Only detections above a certain confidence threshold are considered for further processing to reduce false positives.

After detecting human presence, the system extracts the Region of Interest (ROI) corresponding to each detected individual. This ROI is essentially the portion of the frame containing the detected person. The extraction process ensures that subsequent face detection and recognition steps are focused only on relevant areas, thereby improving efficiency and reducing computational overhead.

The extracted ROI is then passed to the face detection and recognition module. Face detection is performed using the DeepFace framework, which internally uses advanced models such as RetinaFace to accurately locate faces within an image. Once a face is detected, it is processed to generate a numerical representation known as a facial embedding. This embedding captures unique facial features and is used for identity comparison.

The generated embedding is compared with a pre-existing database of known individuals. The system calculates similarity scores between the input embedding and stored embeddings. If the similarity exceeds a predefined threshold,

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 6, Issue 14, April 2026
Impact Factor: 8.2

the individual is classified as a known person; otherwise, the individual is marked as unknown. This classification step is crucial for determining whether the system should trigger security actions.

In the case of known individuals, the system simply logs the event along with details such as name, timestamp, and confidence score. This information is stored in the database for future reference and analysis. For unknown individuals, the system performs additional actions as part of the intruder handling mechanism. It captures the image of the detected face and stores it for record-keeping. Additionally, a sketch-like representation of the face is generated using image processing techniques such as edge detection and grayscale transformations. This feature enhances visibility in scenarios where the captured image is blurred or unclear.

The system incorporates a multi-tenant database architecture to manage event logs. This architecture allows different organizations to use separate databases while sharing the same application codebase. The database configuration is dynamically selected using a configuration file, enabling support for various database systems such as SQLite, MySQL, and PostgreSQL. This approach enhances scalability and flexibility, making the system suitable for deployment across multiple institutions.

To ensure real-time awareness of security threats, the system includes an alert module that sends email notifications whenever an unknown individual is detected. The email alert is implemented using the Simple Mail Transfer Protocol (SMTP) and contains relevant details such as timestamp and intruder status. This feature ensures that administrators are immediately informed and can take necessary actions without delay.

In addition to local processing and alerting, the system provides an API layer using FastAPI. This layer exposes RESTful endpoints that allow external applications or dashboards to access surveillance data. Through these APIs, users can retrieve event logs, monitor system activity, and analyze performance metrics in real time. The inclusion of an API layer enhances the interoperability of the system and supports integration with modern web-based platforms.

Overall, the methodology combines multiple advanced technologies into a cohesive pipeline that ensures efficient and intelligent surveillance. The use of deep learning models for detection and recognition, combined with scalable database architecture and real-time alerting, enables the system to operate effectively in dynamic environments. The modular design also allows for future enhancements, making the system adaptable to evolving security requirements.

IV. SYSTEM ARCHITECTURE

The proposed smart surveillance system follows a modular and layered architecture designed to ensure scalability, flexibility, and real-time performance. The system is composed of multiple interconnected modules, each responsible for a specific task in the surveillance pipeline. These modules include the input layer, detection layer, recognition layer, decision layer, database layer, alert system, and API layer.

The architecture begins with the input layer, where video data is captured from various sources such as webcams, CCTV cameras, or IP camera streams. This flexibility allows the system to be deployed across different environments without requiring significant hardware changes. The captured video is processed frame by frame to maintain continuous monitoring.

The next stage is the detection layer, which utilizes the YOLOv8 model to identify human presence within each frame. YOLO processes the entire image efficiently and outputs bounding boxes around detected individuals along with confidence scores. Only detections with sufficient confidence are forwarded to the next stage to ensure accuracy and reduce noise.

Following detection, the recognition layer extracts the Region of Interest (ROI) corresponding to each detected person and performs face detection using the DeepFace framework. Once a face is detected, facial embeddings are generated and compared with stored embeddings in the database. This comparison enables the system to identify individuals as either known or unknown.

The decision layer plays a critical role in determining the system's response. If a match is found, the individual is labeled as known, and the event is logged accordingly. If no match is found, the individual is classified as an intruder.

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 6, Issue 14, April 2026
Impact Factor: 8.2

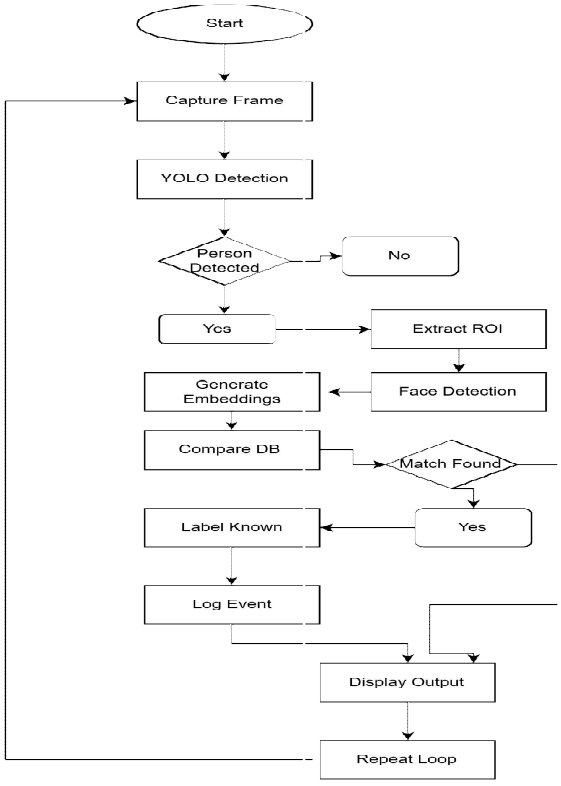
In such cases, the system triggers additional actions, including capturing the intruder's image and generating a sketch-like representation to improve visibility in low-quality conditions.

The database layer is implemented using a multi-tenant architecture, which allows multiple organizations to use the same system with different database configurations. Depending on the configuration file, the system can connect to databases such as SQLite, MySQL, or PostgreSQL. This approach enhances scalability and makes the system adaptable to different deployment scenarios.

The alert system is responsible for notifying administrators in real time. When an intruder is detected, an email alert is sent using SMTP, ensuring immediate awareness of potential threats. Additionally, all events are stored in the database for future analysis and auditing.

Finally, the API layer, implemented using FastAPI, exposes RESTful endpoints that allow external systems to access surveillance data. This enables integration with dashboards, monitoring tools, and other applications, thereby improving usability and extensibility.

Overall, the modular design of the system ensures efficient processing, easy integration, and scalability, making it suitable for real-world surveillance applications.



V. IMPLEMENTATION

The proposed smart surveillance system is implemented using Python due to its extensive support for computer vision and machine learning libraries. The core detection module is built using YOLOv8 from the Ultralytics framework, which enables real-time human detection with high accuracy and efficiency. The model processes live video frames and identifies human presence by generating bounding boxes along with confidence scores.

For facial recognition, the DeepFace library is utilized. It provides a high-level interface for face detection and embedding generation using pre-trained deep learning models. Faces are extracted from the detected regions of interest

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 6, Issue 14, April 2026 Impact Factor: 8.2

(ROI), and embeddings are generated and compared against a stored database to identify individuals. A custom recognition module is implemented to handle this comparison and classification process.

OpenCV is used extensively for video capture, frame processing, ROI extraction, and image manipulation. It also plays a role in generating sketch-like representations of detected intruders using basic image processing techniques such as grayscale conversion and edge detection.

The backend system is designed using FastAPI, which provides RESTful API endpoints for accessing surveillance data such as event logs. This enables integration with external dashboards and monitoring systems. The database layer is implemented using a multi-tenant architecture, where configuration files determine the database type, allowing support for SQLite, MySQL, and PostgreSQL.

Additionally, an email alert system is implemented using the SMTP protocol. When an unknown individual is detected, the system automatically sends an alert email to the administrator. The entire system is modular, ensuring ease of maintenance and future scalability.

VI. RESULTS AND DISCUSSION

The proposed smart surveillance system was tested in a real-time environment using a webcam as well as recorded video inputs to evaluate its performance in detection, recognition, and alert generation. The YOLOv8-based detection module demonstrated high accuracy in identifying human presence, even in moderately crowded scenes, while maintaining real-time processing speed. The system was able to consistently detect individuals with reliable bounding box localization and confidence scores.

The facial recognition module, implemented using DeepFace, showed effective performance in identifying known individuals when sufficient facial data was available. The system correctly recognized registered users and labeled them accordingly with high confidence. In cases where an unregistered individual appeared, the system classified the person as “unknown” and triggered the intruder detection workflow. This included capturing the image, generating a sketch-like representation, logging the event in the database, and sending an email alert to the administrator.

The multi-tenant database architecture functioned as expected, allowing seamless switching between different database configurations without affecting system performance. Event logs, including timestamps, identity status, and confidence scores, were stored accurately and could be retrieved using API endpoints. The FastAPI-based backend provided smooth access to system data, enabling real-time monitoring and integration with external interfaces.

However, certain limitations were observed during testing. The system’s performance was affected under low lighting conditions and when faces were partially occluded. Additionally, recognition accuracy depended on the quality and quantity of stored facial data. Despite these limitations, the overall system demonstrated reliable performance and achieved the objective of automating surveillance tasks. The integration of detection, recognition, alerting, and data management into a unified framework highlights the effectiveness of the proposed solution in real-world applications.

VII. CONCLUSION

In this paper, an AI-based smart surveillance system has been presented, integrating real-time human detection, facial recognition, and automated alert mechanisms into a unified framework. The system leverages advanced deep learning models such as YOLOv8 for efficient object detection and DeepFace for accurate identity recognition. By combining these technologies, the system is capable of continuously monitoring video streams, identifying individuals, and distinguishing between known and unknown persons in real time.

A key contribution of the proposed system is the implementation of a multi-tenant database architecture, which enhances scalability and flexibility by allowing multiple organizations to operate independently within the same system. The integration of additional features such as image capture, sketch generation, and email-based alerting further strengthens the system’s ability to respond effectively to potential security threats. Moreover, the inclusion of a FastAPI-based backend enables seamless data access and integration with external applications.

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 6, Issue 14, April 2026
Impact Factor: 8.2

The experimental results demonstrate that the system performs reliably in real-time conditions, achieving accurate detection and recognition while maintaining efficient processing speed. Although certain limitations exist, such as sensitivity to lighting conditions and dependency on image quality, the overall system successfully achieves its objective of automating surveillance and reducing human intervention. The proposed solution provides a practical and scalable approach to modern security challenges and can be effectively deployed in various real-world environments.

REFERENCES

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv preprint arXiv:2004.10934*, 2020.

[3] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," *Ultralytics*, 2023. [Online]. Available: https://docs.ultralytics.com

[4] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[5] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," *British Machine Vision Conference (BMVC)*, 2015.

[6] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015.

[7] A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

[8] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015.

[9] A. Rosebrock, "OpenCV: Open Source Computer Vision Library," *PyImageSearch*, 2015.

[10] S. Ramírez et al., "FastAPI: High Performance Web Framework for Building APIs with Python," *Journal of Open Source Software*, 2020.

[11] I. Goodfellow et al., "Generative Adversarial Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[12] R. Szeliski, "Computer Vision: Algorithms and Applications," Springer, 2010.