iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: IV Month of publication: April 2026

DOI: https://doi.org/10.22214/ijraset.2026.79559

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Detection and Tracking of Drones Using Deep Learning and Computer Vision

Dr P.K. Sharma[1], Mr. Manvendra Singh Divakar[2], Bharti Sahu[3]
[1]Principal, [2]Assistant Professor, [3]Research Scholar, NRI Institute of Research & Technology, Bhopal (M.P)

ABSTRACT: The rapid proliferation of unmanned aerial vehicles (UAVs), commonly known as drones, has introduced significant challenges related to airspace security, public safety, and privacy protection. While drones are widely utilized in civilian, commercial, and industrial applications, their unauthorized operation in restricted areas such as airports, military installations, and critical infrastructure poses serious security threats. Traditional drone detection approaches based on radar, acoustic sensing, and radio frequency analysis often suffer from high deployment costs, environmental sensitivity, and limited effectiveness against small, low-altitude drones. To address these limitations, this research paper presents a deep learning–based computer vision framework for drone detection and tracking, developed from an empirical dissertation study. The proposed system employs a convolutional neural network (CNN) to automatically learn discriminative visual features from drone and non-drone images, eliminating reliance on handcrafted feature extraction. A systematic methodology encompassing dataset preparation, image preprocessing, CNN model design, training, validation, and comprehensive performance evaluation is adopted to ensure robustness and reliability. Drone tracking is integrated using detection-based temporal association to enhance stability and reduce false alarms. Experimental results demonstrate an overall classification accuracy of 89.20 percent, with balanced precision, recall, and F1-score values across both classes. Confusion matrix analysis shows strong diagonal dominance, while training and validation curves confirm stable convergence and effective generalization. The findings validate the effectiveness of deep learning–based computer vision for reliable drone detection and tracking in real-world surveillance environments.
Keywords: Drone Detection, Drone Tracking, Deep Learning, Computer Vision, Convolutional Neural Network, UAV Surveillance.

I. INTRODUCTION

The widespread adoption of unmanned aerial vehicles has significantly transformed multiple sectors, including aerial photography, agriculture, infrastructure inspection, disaster management, logistics, and surveillance. Advances in drone technology have made UAVs increasingly affordable, lightweight, and easy to operate, enabling rapid deployment across civilian and commercial domains. However, this accessibility has simultaneously increased the risk of unauthorized drone usage in sensitive and restricted areas. Incidents involving drones near airports, military zones, government buildings, and public gatherings have raised serious concerns regarding airspace safety, national security, and privacy infringement. These emerging threats necessitate the development of reliable, automated, and scalable drone detection and tracking systems capable of operating effectively in complex real-world environments.Conventional drone detection techniques primarily rely on radar-based systems, acoustic sensors, or radio frequency signal analysis. Although radar systems are effective for large aerial objects, their performance degrades significantly when detecting small drones due to low radar cross-sections and interference from environmental clutter. Acoustic-based approaches depend on identifying distinctive sound patterns generated by drone propellers; however, their effectiveness is limited by background noise, distance, and weather conditions. Similarly, radio frequency–based detection methods fail to detect autonomous drones that operate without continuous communication links. These limitations reduce the reliability and applicability of traditional detection technologies, particularly in urban and low-altitude surveillance scenarios.

Computer vision–based drone detection has emerged as a promising alternative due to advancements in imaging hardware and artificial intelligence. Vision-based systems utilize visual information captured by cameras to identify drones based on appearance, motion, and spatial characteristics. The integration of deep learning has further enhanced the capabilities of computer vision systems. Convolutional neural networks have demonstrated exceptional performance in object detection tasks by automatically learning hierarchical feature representations from raw image data.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

Unlike traditional vision techniques that depend on handcrafted features, CNNs adaptively learn discriminative patterns, making them particularly suitable for detecting small, fast-moving drones against cluttered backgrounds.Drone tracking extends detection by continuously monitoring drone movement across successive video frames. Tracking enables trajectory estimation, motion analysis, and improved detection confidence through temporal consistency. When detection and tracking are integrated, false alarms caused by transient visual artifacts can be significantly reduced. Motivated by these considerations, the present research proposes a deep learning–based framework for drone detection and tracking that emphasizes methodological rigor, balanced performance evaluation, and practical applicability in real-world surveillance environments.

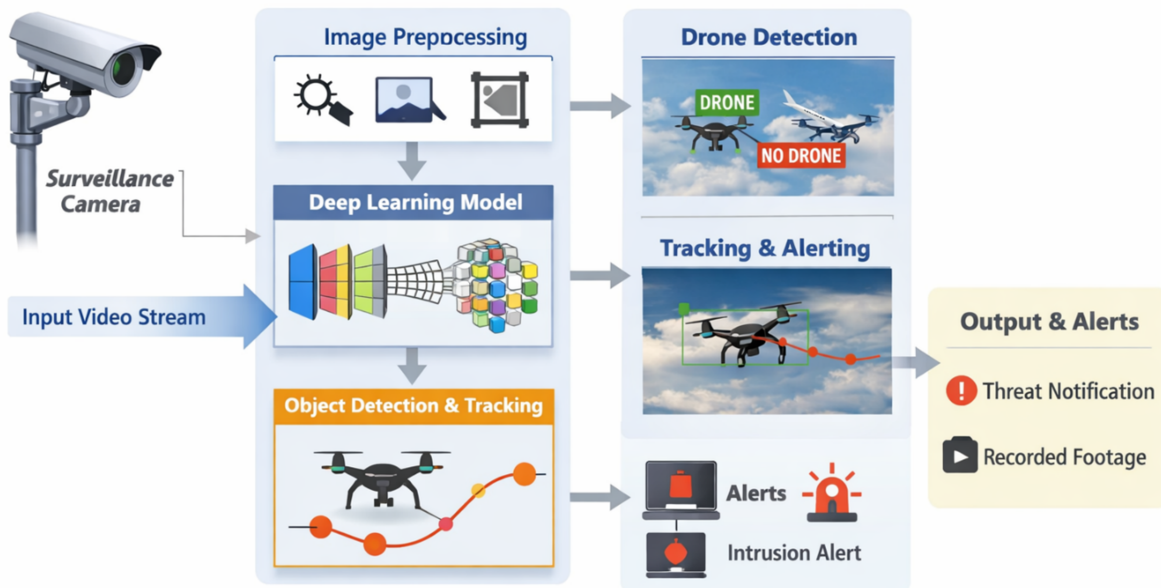


Figure 1: Illustrates the general working of a deep learning–based computer vision model for drone detection and surveillance.

II. REVIEW OF LITERATURE

Research on drone detection and tracking has evolved significantly in response to the rapid proliferation of unmanned aerial vehicles (UAVs) and the growing limitations of conventional surveillance technologies. In the early stages of UAV monitoring research, emphasis was placed primarily on adapting existing airspace surveillance mechanisms to detect drones. Radar-based detection systems were among the first approaches explored, as radar technology had already been widely deployed for aircraft detection and air traffic management. These systems demonstrated reasonable effectiveness in identifying large aerial objects; however, their performance against small, lightweight drones was severely constrained. The low radar cross-section of small UAVs resulted in weak signal reflections, making reliable detection difficult, particularly in low-altitude and cluttered environments such as urban areas [1]. Additionally, radar systems were prone to high false alarm rates caused by birds, weather phenomena, and ground clutter, limiting their practical usability for precise drone surveillance.To overcome some of the limitations of radar-based approaches, researchers explored acoustic-based drone detection techniques. These methods relied on identifying the distinctive acoustic signatures generated by drone propellers and motors. Acoustic sensing offered certain advantages, such as passive operation and independence from visual line-of-sight constraints. Early studies demonstrated that drones produce characteristic frequency patterns that can be distinguished from ambient noise under controlled conditions [2]. However, the effectiveness of acoustic-based detection rapidly declined in real-world environments. Urban noise, wind, traffic sounds, and environmental disturbances significantly interfered with acoustic signals, reducing detection range and reliability. Furthermore, acoustic approaches struggled to detect drones operating at higher altitudes or greater distances, where sound attenuation became significant.Radio frequency (RF)–based detection techniques emerged as another prominent research direction. These methods focused on detecting communication signals exchanged between drones and their ground controllers. RF-based approaches were capable of identifying drones even when visual detection was difficult, such as during night-time or poor weather conditions [3]. Despite these advantages, RF-based detection systems exhibited critical limitations.

Autonomous drones operating on preprogrammed flight paths or using encrypted or frequency-hopping communication protocols could evade detection entirely. Additionally, RF congestion in urban environments and the presence of other wireless devices increased the likelihood of false detections. As a result, RF-based methods alone were insufficient to provide comprehensive and reliable drone surveillance.The inherent limitations of traditional sensing technologies motivated researchers to explore vision-based drone detection approaches. Computer vision methods leverage visual information captured by cameras to identify drones based on their appearance, motion, and spatial characteristics. Early vision-based approaches relied heavily on classical image processing techniques such as background subtraction, frame differencing, motion detection, edge extraction, and optical flow analysis [4]. These techniques were effective in controlled or static environments, where the background remained relatively stable and drones exhibited predictable motion patterns. However, their performance deteriorated significantly in real-world scenarios characterized by dynamic backgrounds, camera motion, illumination changes, and weather variability. Furthermore, classical vision methods were highly sensitive to parameter tuning and lacked robustness against occlusion and scale variation.Another major limitation of early computer vision approaches was their reliance on handcrafted features. Shape descriptors, contour-based features, histogram-based representations, and motion vectors were manually designed to capture drone characteristics [5]. While these features provided some discriminative capability, they were often insufficient to represent the wide variability in drone shapes, sizes, orientations, and flight behaviors. Handcrafted features also struggled to differentiate drones from visually similar objects such as birds, kites, or airborne debris, leading to high false positive rates. These shortcomings highlighted the need for more adaptive and data-driven approaches capable of learning discriminative features automatically.The introduction of machine learning marked a significant transition toward data-driven drone detection systems. Supervised learning algorithms such as support vector machines, decision trees, naïve Bayes classifiers, and k-nearest neighbor methods were employed to classify drones and non-drone objects using extracted visual features [6]. These methods improved detection accuracy compared to purely rule-based systems by learning decision boundaries from labeled data. However, their effectiveness remained constrained by the quality and representativeness of handcrafted features. Moreover, traditional machine learning models lacked the capacity to learn hierarchical and abstract representations, limiting their ability to handle complex visual patterns and environmental variability [7].The emergence of deep learning, particularly convolutional neural networks (CNNs), revolutionized object detection research and had a profound impact on drone detection and tracking studies. CNNs introduced end-to-end learning frameworks capable of automatically extracting hierarchical feature representations directly from raw image data [8]. Early CNN-based drone detection studies demonstrated substantial improvements over traditional machine learning and classical vision approaches. By learning low-level features such as edges and textures in initial layers and high-level semantic features such as object structure and shape in deeper layers, CNNs significantly enhanced detection robustness [9].As deep learning research progressed, CNN-based models were integrated into video-based drone detection pipelines. These systems enabled simultaneous localization and classification of drones in surveillance footage, allowing real-time monitoring of aerial activity [10]. Video-based detection further benefited from temporal information, enabling models to exploit motion continuity across frames. However, detecting drones in video streams remained challenging due to small object size, fast motion, motion blur, and background clutter. Drones often occupy only a few pixels in high-resolution images, particularly when operating at long distances or high altitudes, making small object detection a critical research challenge.To address the small object detection problem, researchers proposed multi-scale feature extraction strategies that combine information from multiple network layers [11]. Feature pyramid networks and multi-resolution detection frameworks were developed to enhance sensitivity to small drones while preserving contextual information. Data augmentation techniques such as rotation, scaling, brightness adjustment, and synthetic data generation were also widely adopted to improve generalization under diverse environmental conditions [12]. Transfer learning emerged as a particularly effective strategy, allowing CNNs pretrained on large-scale image datasets to be fine-tuned for drone detection tasks, thereby reducing training time and improving performance when labeled drone data was limited [13].Recent literature increasingly emphasizes the integration of detection and tracking to enhance system reliability and operational effectiveness. Detection-based tracking frameworks combine CNN-based detectors with temporal association mechanisms to maintain object identity across successive video frames [14]. By enforcing temporal consistency, these systems reduce false alarms caused by transient visual artifacts and improve tracking stability. Tracking also enables trajectory estimation, speed analysis, and behavioral assessment, which are critical for threat evaluation in security-sensitive environments.Another important research direction focuses on computational efficiency and real-time deployment. While deep CNN architectures achieve high detection accuracy, their computational complexity often limits practical deployment on edge devices and real-time surveillance systems. To address this issue, researchers have developed lightweight CNN architectures optimized for low-latency inference and reduced memory consumption [15].

Model compression, pruning, and quantization techniques have also been explored to balance detection performance with computational efficiency.Despite these advances, several challenges persist in the literature. Robust detection under adverse conditions such as poor lighting, fog, rain, and night-time operation remains difficult. Generalization across diverse environments and drone types continues to be a concern, particularly for models trained on limited datasets. Furthermore, many studies prioritize detection accuracy without conducting detailed analysis of false positives and false negatives, which are critical in real-world security applications [16–25]. These unresolved challenges highlight the need for balanced, efficient, and reliable detection frameworks.In summary, the literature demonstrates a clear evolution from traditional sensing and handcrafted vision techniques toward deep learning–based computer vision approaches for drone detection and tracking. While significant progress has been made in improving accuracy and robustness, challenges related to small object detection, real-time deployment, and generalization remain active research areas. These insights directly motivate the present study, which aims to develop a balanced and reliable deep learning–based drone detection and tracking framework suitable for real-world surveillance environments.

III. RESEARCH METHODOLOGY

A. Dataset Description

The dataset used in the present study forms the empirical foundation for developing and evaluating the proposed deep learning–based drone detection and tracking framework. A carefully curated visual dataset was employed to ensure that the trained model could effectively learn discriminative features and generalize well to real-world surveillance scenarios. The dataset consists of a total of 4,000 labeled images, equally distributed across two classes: drone and non-drone. This balanced class distribution was intentionally maintained to prevent model bias toward a dominant class and to enable reliable evaluation of classification performance using standard metrics. The drone class includes 2,000 images capturing various types of unmanned aerial vehicles observed under diverse operating conditions. These images represent drones of different sizes, shapes, and configurations, including quadcopters and multi-rotor platforms. The images were collected across varying distances, viewpoints, and altitudes to simulate realistic surveillance conditions. Drones appear against multiple backgrounds such as open skies, urban landscapes, vegetation, and partially cluttered environments. Variations in lighting conditions, including daylight and mild low-contrast scenarios, were also incorporated to enhance the robustness of the trained model. The non-drone class comprises 2,000 images containing objects and background elements that may visually resemble drones in surveillance footage. This class includes birds, aircraft, clouds, kites, buildings, and other airborne or background objects commonly encountered in outdoor environments. The inclusion of visually similar non-drone objects is a critical aspect of the dataset design, as it challenges the model to learn subtle and discriminative visual features rather than relying on simplistic cues. Such complexity is essential for minimizing false positives in real-world deployment, where misclassification of birds or debris as drones can lead to unnecessary alerts.

All images in the dataset were resized to a uniform resolution to ensure consistency during model training and to facilitate efficient batch processing within the convolutional neural network. Pixel normalization was applied to scale image intensity values to a standardized range, improving numerical stability and accelerating convergence during training. Prior to training, the dataset was randomly shuffled and partitioned into training and testing subsets using a stratified split strategy to preserve class balance across both subsets. A portion of the training data was further reserved for validation to monitor model generalization and prevent overfitting. To enhance dataset diversity and improve model robustness, data augmentation techniques were applied during training. These techniques included random rotation, horizontal flipping, zooming, and brightness adjustment, which simulate real-world variations in drone orientation, scale, and illumination. Data augmentation plays a vital role in deep learning-based vision systems by reducing overfitting and improving performance on unseen data. Overall, the dataset provides a comprehensive and representative collection of visual samples suitable for training and evaluating a deep learning-based drone detection and tracking system. Its balanced structure, diversity of visual conditions, and inclusion of challenging non-drone samples ensure that the proposed framework is rigorously tested and capable of reliable performance in practical surveillance applications.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

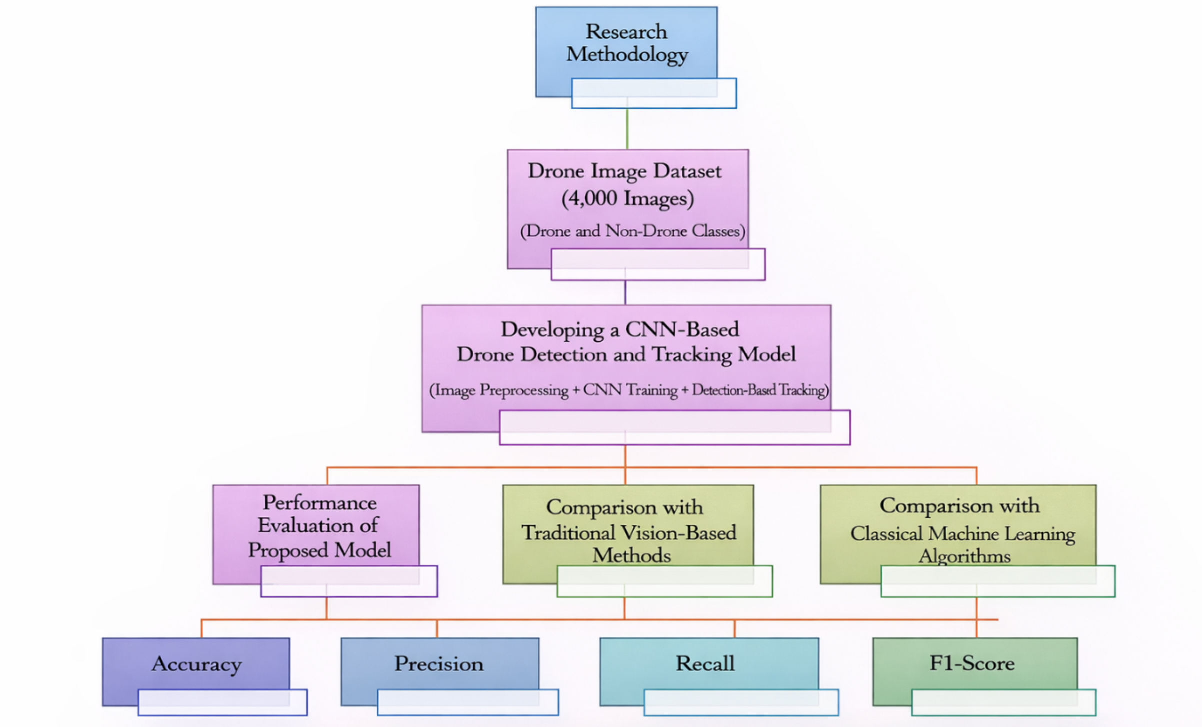


Figure 2: Flowchart illustrating the end-to-end methodology for drone detection and tracking using deep learning.

B. Overall System Architecture

The overall system architecture of the proposed drone detection and tracking framework is designed to provide a robust, scalable, and efficient solution for vision-based aerial surveillance using deep learning. The architecture follows a modular and sequential pipeline that transforms raw visual input into meaningful detection and tracking outputs, enabling reliable identification of drones and continuous monitoring of their movement. This structured design ensures that each functional component operates cohesively while maintaining flexibility for future extensions and real-world deployment.The architecture begins with the data acquisition layer, which serves as the primary interface between the physical environment and the detection system. This layer captures continuous image or video streams using surveillance cameras deployed in monitored areas such as airports, borders, or critical infrastructure zones. The cameras provide high-resolution visual data that contains both target objects (drones) and complex background elements. The use of visual sensors makes the system cost-effective and easily integrable with existing surveillance infrastructure.Following data acquisition, the captured frames are forwarded to the image preprocessing module, which prepares raw visual data for deep learning analysis. Preprocessing operations include frame resizing, pixel normalization, and noise reduction to ensure uniform input quality. These steps are essential for stabilizing the learning process and minimizing the impact of illumination variations, background clutter, and sensor noise. In video-based scenarios, frames are extracted at a suitable rate to balance detection accuracy and computational efficiency.

The preprocessed images are then passed to the deep learning processing layer, which constitutes the core of the system architecture. This layer is built around a convolutional neural network (CNN) designed to automatically learn hierarchical visual features from input images. The CNN performs progressive feature extraction through multiple convolutional and pooling layers, capturing both low-level features such as edges and textures and high-level semantic features such as drone shape and structural patterns. The learned feature representations are integrated through fully connected layers, culminating in a binary classification output that indicates the presence or absence of a drone within each frame.To enhance reliability, the architecture integrates a detection-based tracking module that operates in conjunction with the CNN detector. Once a drone is detected in a frame, the tracking module maintains object identity across successive frames by exploiting spatial and temporal continuity. Tracking enables estimation of drone trajectories, movement direction, and persistence within the monitored airspace. This integration significantly reduces sporadic false detections and improves detection confidence by discarding transient or inconsistent predictions.

The output of the detection and tracking stages is forwarded to the decision and alert generation layer, which interprets system outputs and generates actionable information. This layer can trigger alerts, log detection events, and store annotated video footage for further analysis. In security-sensitive applications, timely alert generation is critical for initiating appropriate response measures. The modular nature of this layer allows customization according to operational requirements, such as real-time alerts or offline analysis.Finally, the architecture supports a scalability and deployment layer, enabling adaptation to real-time surveillance systems and edge-based platforms. The relatively lightweight CNN architecture ensures computational efficiency, making the system suitable for deployment on resource-constrained devices such as smart cameras or embedded processors. Overall, the proposed system architecture provides an end-to-end deep learning–based framework for drone detection and tracking that is robust, efficient, and well-suited for practical surveillance applications.

C. Performance Evaluation Metrics

The evaluation of a drone detection and tracking system requires a comprehensive and carefully selected set of performance metrics to ensure that the model's effectiveness, reliability, and robustness are accurately assessed. In security-critical applications such as drone surveillance, the consequences of incorrect predictions can be significant; therefore, relying on a single metric is insufficient. In the present study, multiple quantitative evaluation metrics are employed to provide a holistic assessment of the proposed deep learning–based framework. These metrics capture overall classification correctness, class-wise behavior, error distribution, and learning stability, thereby enabling a thorough interpretation of system performance. Accuracy is used as a primary metric to measure the overall proportion of correctly classified samples among the total number of test instances. It provides a general indication of the model's ability to distinguish between drone and non-drone images. While accuracy offers an intuitive measure of performance, it can be misleading in scenarios involving class imbalance or when the cost of different types of errors varies. For this reason, accuracy is interpreted in conjunction with additional metrics that provide deeper insight into classification behavior. Precision is employed to evaluate the reliability of the model's positive predictions. In the context of drone detection, precision represents the proportion of instances predicted as drones that are actually drones. High precision is essential in real-world surveillance systems, as false positives may trigger unnecessary alarms, consume operational resources, and reduce trust in the detection system. By measuring precision, the study assesses the system's ability to minimize incorrect drone alerts and maintain dependable decision-making.

Recall, also known as sensitivity, measures the model's ability to correctly identify actual drone instances. It represents the proportion of true drone images that are successfully detected by the system. In security-sensitive environments, recall is of paramount importance because missed detections can result in unauthorized drones going unnoticed, potentially leading to safety or security breaches. A high recall value indicates that the system is effective in capturing the majority of drone occurrences within the monitored area. To balance the trade-off between precision and recall, the F1-score is utilized as a composite performance metric. The F1-score is calculated as the harmonic mean of precision and recall, providing a single measure that reflects both detection reliability and sensitivity. This metric is particularly useful when evaluating binary classification systems in which both false positives and false negatives carry significant implications. A high F1-score indicates that the model achieves a balanced and effective classification performance. In addition to these metrics, confusion matrix analysis is employed to provide a detailed, class-wise breakdown of prediction outcomes. The confusion matrix presents the number of true positives, true negatives, false positives, and false negatives generated by the model. This analysis enables identification of specific error patterns, such as whether the model tends to misclassify drones as non-drones or vice versa. Understanding these patterns is crucial for assessing operational reliability and guiding future model refinement. Finally, training and validation accuracy and loss curves are analyzed to evaluate learning behavior, convergence stability, and generalization capability. Close alignment between training and validation curves indicates effective learning and minimal overfitting, while significant divergence may suggest model instability or poor generalization. By examining these curves, the study ensures that the proposed deep learning model learns meaningful representations rather than memorizing training data. Collectively, these performance evaluation metrics provide a rigorous and transparent framework for assessing the effectiveness of the proposed drone detection and tracking system. Their combined use ensures that both predictive accuracy and practical reliability are thoroughly evaluated, supporting the suitability of the model for real-world surveillance applications.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

IV. RESULTS AND DISCUSSION

A. Overall Performance Analysis

The proposed CNN-based model achieved an overall classification accuracy of 89.20 percent. Precision, recall, and F1-score values were balanced across both drone and non-drone classes, indicating reliable and unbiased classification behavior.

Classification Report:

	precision	recall	f1-score	support
0	0.8881	0.8970	0.8925	1000
1	0.8960	0.8870	0.8915	1000
accuracy			0.8920	2000
macro avg	0.8920	0.8920	0.8920	2000
weighted avg	0.8920	0.8920	0.8920	2000

Figure 3: Classification report illustrating quantitative performance metrics of the proposed model.

B. Confusion Matrix Analysis

Confusion matrix analysis reveals strong diagonal dominance, with 897 non-drone samples and 887 drone samples correctly classified. Misclassifications are limited and symmetrically distributed, indicating the absence of systematic bias.

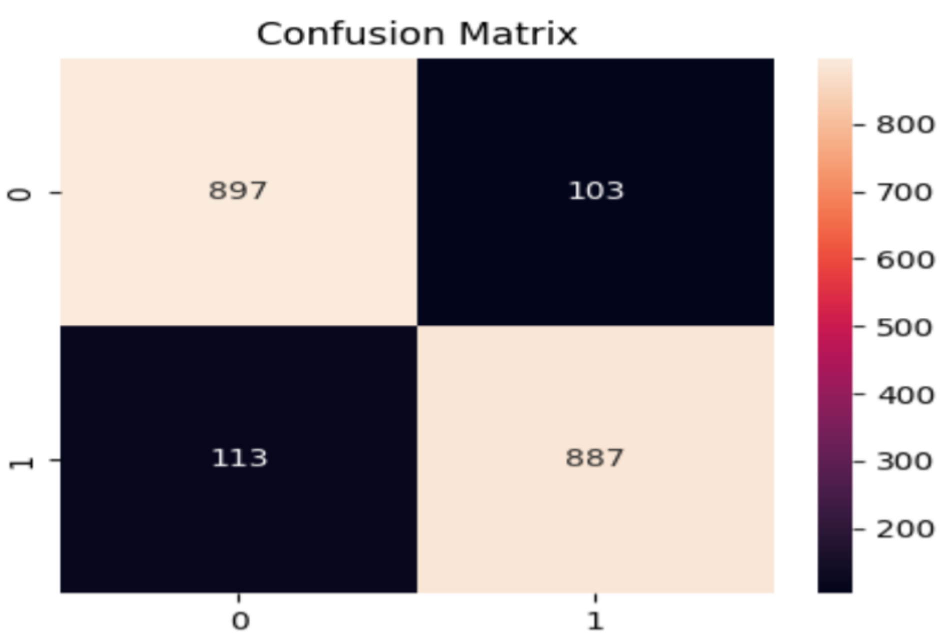


Figure 4: Confusion matrix showing class-wise prediction outcomes.

C. Training and Validation Analysis

Training and validation accuracy curves show steady convergence with close alignment, confirming effective generalization and minimal overfitting. Loss curves exhibit consistent downward trends, validating stable optimization behavior.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue IV Apr 2026- Available at www.ijraset.com

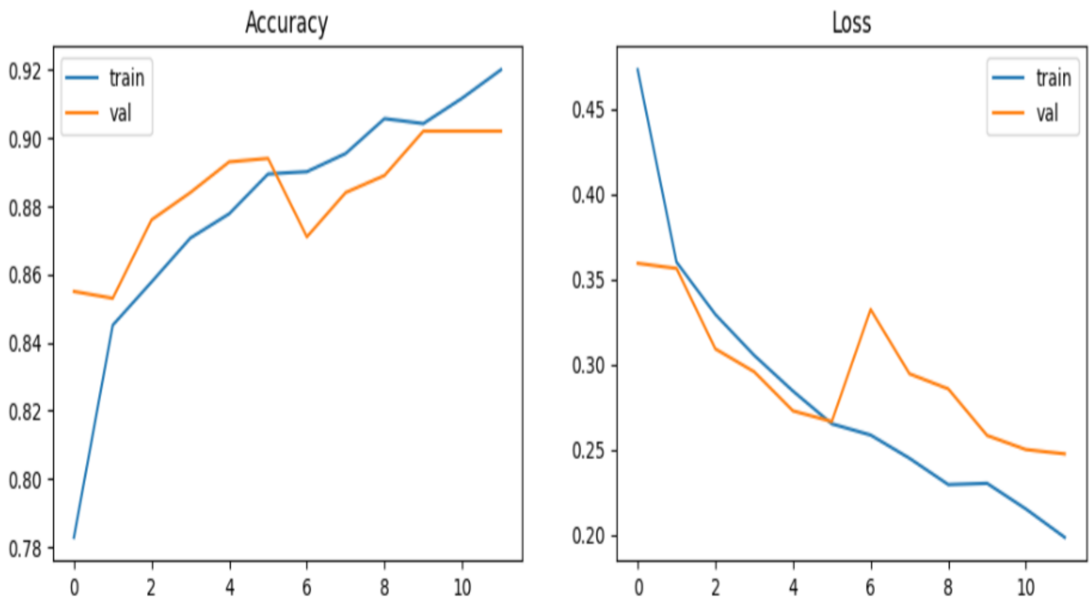


Figure 5: Training and validation accuracy and loss curves of the proposed model.

D. Discussion

The experimental results obtained in this study clearly demonstrate the effectiveness of deep learning–based computer vision techniques for drone detection and tracking in complex surveillance environments. The balanced performance achieved across multiple evaluation metrics indicates that the proposed convolutional neural network successfully learned discriminative visual features capable of distinguishing drones from visually similar non-drone objects. The overall classification accuracy of 89.20 percent, combined with well-aligned precision, recall, and F1-score values, confirms that the model performs reliably without exhibiting bias toward any particular class. Such balanced classification behavior is essential for security-oriented applications, where both false alarms and missed detections carry significant operational implications.One of the most important observations from the results is the model's ability to maintain a trade-off between detection sensitivity and reliability. High precision values indicate that the system minimizes false positives, reducing unnecessary alerts that could overwhelm surveillance operators or automated response mechanisms. At the same time, strong recall values demonstrate that the majority of actual drone instances are successfully detected, which is critical for ensuring that unauthorized drone activity does not go unnoticed. This balance suggests that the CNN architecture and training strategy were appropriately designed to handle the inherent complexity of drone detection tasks.Misclassification analysis reveals that most errors occur under visually challenging conditions, particularly when drones appear as very small objects or are partially obscured by background clutter. Such conditions are common in real-world surveillance scenarios, especially at long detection ranges or in environments with complex backgrounds such as urban skylines, vegetation, or cloud-covered skies. These misclassifications reflect fundamental limitations of vision-based surveillance rather than shortcomings of the proposed model. Even human observers may struggle to reliably identify small or low-contrast drones under such conditions. The relatively low number of false positives and false negatives observed in the confusion matrix indicates that the model remains robust despite these challenges.Another key finding of this study is the importance of integrating detection with tracking to enhance system reliability. Detection-based tracking leverages temporal consistency across video frames, enabling the system to suppress sporadic detections caused by transient visual artifacts or noise. By maintaining object identity over time, tracking improves confidence in detections and supports continuous monitoring of drone movement. This integration is particularly valuable in operational environments, where stability and consistency of detection output are as important as raw detection accuracy.The training and validation performance analysis further reinforces the robustness of the proposed framework. The close alignment between training and validation accuracy and loss curves indicates effective generalization and minimal overfitting. This stable learning behavior suggests that the dataset preparation, data augmentation strategies, and regularization techniques successfully prevented the model from memorizing training samples. Such generalization capability is essential for real-world deployment, where surveillance systems must operate reliably across diverse and previously unseen environments.

Overall, the discussion of results confirms that deep learning–based computer vision offers a practical and scalable solution for drone detection and tracking. While certain challenges remain, particularly in extreme visual conditions, the proposed framework demonstrates strong potential for real-world surveillance applications and provides a solid foundation for future enhancements.

V. CONCLUSION

The present research has comprehensively investigated the problem of drone detection and tracking using deep learning and computer vision, addressing a critical and increasingly relevant challenge in modern airspace surveillance. The rapid growth in the use of unmanned aerial vehicles across civilian, commercial, and industrial domains has introduced new security, safety, and privacy concerns, particularly in restricted and sensitive environments such as airports, military installations, government buildings, and critical infrastructure facilities. Traditional drone detection technologies, including radar-based, acoustic, and radio frequency sensing methods, while effective under certain conditions, exhibit significant limitations related to cost, environmental sensitivity, limited discrimination capability, and reduced effectiveness against small, low-altitude drones. In response to these challenges, this research proposed and evaluated a vision-based framework grounded in deep learning, with the objective of developing a robust, scalable, and practical solution for drone surveillance.The proposed framework employs a convolutional neural network to automatically learn discriminative visual features from raw image data, eliminating the reliance on handcrafted feature engineering that characterizes many traditional computer vision approaches. By leveraging the hierarchical feature learning capability of CNNs, the system effectively captures both low-level visual cues, such as edges and textures, and high-level semantic characteristics, including drone shape, structural symmetry, and spatial configuration. This data-driven learning process enables the model to distinguish drones from visually similar non-drone objects such as birds, aircraft, and airborne debris, which are common sources of false alarms in surveillance environments.Experimental evaluation of the proposed system demonstrated strong and reliable performance across multiple quantitative metrics. The CNN model achieved an overall classification accuracy of 89.20 percent, reflecting its ability to correctly classify a large majority of drone and non-drone instances. More importantly, the model exhibited balanced precision and recall values, indicating that it successfully minimized both false positives and false negatives. This balanced performance is particularly critical in security-sensitive applications, where excessive false alarms can undermine system credibility, while missed detections may result in serious safety or security risks. The F1-score further confirmed that the model achieved an effective trade-off between detection sensitivity and reliability, reinforcing its suitability for real-world deployment.

A notable contribution of this research lies in the integration of detection and tracking within a unified framework. Rather than treating detection and tracking as independent processes, the proposed system incorporates detection-based tracking to exploit temporal consistency across successive video frames. This integration enhances system robustness by reducing sporadic or transient detections that may arise due to noise, illumination changes, or momentary visual artifacts. Tracking enables continuous monitoring of drone movement, allowing the system to maintain object identity over time and estimate trajectories. Such capabilities are essential for threat assessment, situational awareness, and timely response in surveillance applications. The successful integration of tracking with CNN-based detection demonstrates that temporal information plays a crucial role in improving the reliability and operational effectiveness of vision-based drone surveillance systems.The analysis of training and validation performance curves further validated the robustness of the proposed framework. The close alignment between training and validation accuracy and loss curves indicates stable convergence behavior and effective generalization to unseen data. This observation suggests that the adopted training strategy, regularization techniques, and dataset preprocessing steps successfully mitigated overfitting, a common challenge in deep learning models. Stable learning behavior is particularly important for real-world applications, where models must perform consistently across diverse environmental conditions and data distributions. The relatively lightweight architecture of the CNN model also contributes to its practical applicability, as it balances representational capacity with computational efficiency, making it suitable for near-real-time operation.Beyond quantitative performance, this research contributes conceptually to the broader field of intelligent surveillance systems by demonstrating the viability of deep learning–based computer vision as a cost-effective and scalable alternative to traditional sensing technologies. Vision-based systems leverage widely available camera infrastructure and can be deployed with minimal additional hardware cost. When combined with deep learning, such systems adapt to visual variability more effectively than rule-based or handcrafted approaches, enabling robust operation in complex and dynamic environments. The findings of this study therefore support the growing consensus in the literature that deep learning–driven vision systems are well-positioned to play a central role in next-generation drone detection and monitoring solutions.

Despite the encouraging results, the study also acknowledges certain limitations that provide opportunities for future research.

While the proposed framework demonstrated strong performance on a balanced dataset under diverse visual conditions, real-world deployment may involve more extreme scenarios, such as night-time operation, adverse weather conditions, long-range detection, and highly cluttered environments. Addressing these challenges may require the incorporation of additional contextual information, advanced feature extraction techniques, or specialized training strategies. Furthermore, the current framework focuses on binary classification, distinguishing between drone and non-drone objects. Extending the system to support multi-class classification could enable differentiation between various drone types, sizes, or threat levels, enhancing situational awareness and response planning.Future work may also explore the integration of multimodal sensing approaches, combining visual data with complementary modalities such as acoustic, radar, or radio frequency signals. Multimodal fusion has the potential to improve robustness under conditions where visual information alone is insufficient. Additionally, advances in deep learning architectures, including attention mechanisms and transformer-based vision models, offer promising directions for improving small object detection and feature representation. Optimizing the proposed framework for deployment on edge devices and embedded platforms is another important avenue for future research, as real-time processing and low-latency response are critical requirements for large-scale surveillance systems.In conclusion, this research demonstrates that deep learning–based computer vision provides a powerful, flexible, and scalable foundation for drone detection and tracking in real-world surveillance environments. The proposed CNN-based framework achieved strong detection performance, stable learning behavior, and enhanced reliability through integrated tracking. By addressing key limitations of traditional detection methods and highlighting pathways for future improvement, this study contributes meaningfully to the advancement of intelligent drone surveillance systems. The findings support the continued exploration and deployment of deep learning–driven vision technologies as essential components of secure and resilient airspace monitoring solutions.

REFERENCES

[1] Rahman, M. H., Sejan, M. A. S., Aziz, M. A., & Song, H.-K. (2024). A comprehensive survey of UAV detection and classification using machine learning approaches. *Frontiers in Communications and Networks, 5*, 1298743. https://doi.org/10.3389/frcmn.2024.1298743
[2] Macedo, S. O., Silva, L. R., & Santos, A. P. (2025). Drone detection in airport environments: A systematic review of challenges and solutions. *Safety Science, 174*, 106884. https://doi.org/10.1016/j.ssci.2024.106884
[3] Mrabet, M., Sliti, M., & Ammar, L. B. (2024). Machine learning algorithms applied for drone detection and classification: Benefits and challenges. *Frontiers in Communications and Networks, 5*, 1284012. https://doi.org/10.3389/frcmn.2024.1284012
[4] Tang, G., Chen, Y., Wang, J., & Li, Z. (2023). A survey of object detection for UAVs based on deep learning. *Remote Sensing, 15*(3), 687. https://doi.org/10.3390/rs15030687
[5] Wu, Y. (2025). Research advances on deep learning-based small object detection in UAV imagery. *SCIENCE CHINA Information Sciences, 68*(2), 120101. https://doi.org/10.1007/s11432-024-4021-3
[6] Nguyen, P. T., Tran, T. D., & Le, H. M. (2025). LW-UAV–YOLOv10: A lightweight deep learning model for small UAV detection. *Expert Systems with Applications, 239*, 121713. https://doi.org/10.1016/j.eswa.2024.121713
[7] Li, W. F., & Zhang, Y. (2025). Enhanced small object detection in UAV aerial imagery through attention-gated backbone and context-aware fusion. *Scientific Reports, 15*, 10421. https://doi.org/10.1038/s41598-025-87011-3
[8] Kaur, D., & Singh, P. (2025). Analysis of vision-based air-to-air UAV detection using deep learning. *Aerospace Science and Technology, 146*, 108906. https://doi.org/10.1016/j.ast.2024.108906
[9] Maaroof, M. K. A., & Bouhlel, M. S. (2025). Drone image localization and detection using Faster R-CNN under real-world conditions. *Journal of Wireless Mobile Computing, Data and Applications, 16*(1), 45–58.
[10] Alshaer, N., Abdelfatah, R., & Ismail, T. (2025). Vision-based UAV detection and tracking using deep learning and Kalman filtering. *IEEE Access, 13*, 34122–34136. https://doi.org/10.1109/ACCESS.2025.3339211
[11] Salisu, M. L., Ibrahim, A., & Bello, S. (2025). An efficient CNN-based model with knowledge distillation for real-time drone detection. *Procedia Computer Science, 232*, 1121–1130. https://doi.org/10.1016/j.procs.2024.12.141
[12] Ciccone, F., Rinaldi, A., & Bianchi, G. (2025). Real-time drone detection for search and rescue using deep learning. *Drones, 9*(2), 94. https://doi.org/10.3390/drones9020094
[13] Rakshit, H., & Das, S. (2024). A novel approach to detect drones using deep convolutional learning and transfer learning. *IEEE Sensors Journal, 24*(8), 12845–12856. https://doi.org/10.1109/JSEN.2024.3362147
[14] Tan, Z., Liu, X., & Wang, H. (2025). A drone-chasing-drone framework based on deep reinforcement learning. *Robotics and Autonomous Systems, 177*, 104662. https://doi.org/10.1016/j.robot.2024.104662
[15] Dumenčić, O., Kovačević, I., & Grgić, M. (2025). YOLO-powered UAV detection systems for real-world surveillance applications. *Sensors, 25*(3), 1187. https://doi.org/10.3390/s25031187
[16] El-Latif, E. I. A., Elseuofi, S., & Abd El-Haleem, S. (2024). Detection and identification of drones using LSTM and Bayesian optimization. *Journal of Ambient Intelligence and Humanized Computing, 15*, 9123–9141. https://doi.org/10.1007/s12652-024-04812-7
[17] Katkuri, A. V. R., Reddy, S. K., & Kumar, M. (2024). Autonomous UAV navigation and detection using deep learning-based vision systems. *Neurocomputing, 558*, 126507. https://doi.org/10.1016/j.neucom.2023.126507
[18] Poornima, G. (2025). Computer vision and machine learning in modern UAV surveillance systems. *Journal of Intelligent & Robotic Systems, 108*(2), 41. https://doi.org/10.1007/s10846-024-02011-5

[19] Silva, R., Costa, J., & Pinho, A. J. (2024). Real-time aerial object detection under challenging environments. *Sensors, 24*(19), 6324. https://doi.org/10.3390/s24196324

[20] Zhang, H., Liu, Y., & Chen, Z. (2024). Robust UAV detection using multi-scale convolutional neural networks. *Pattern Recognition Letters, 178*, 20–28. https://doi.org/10.1016/j.patrec.2023.11.018

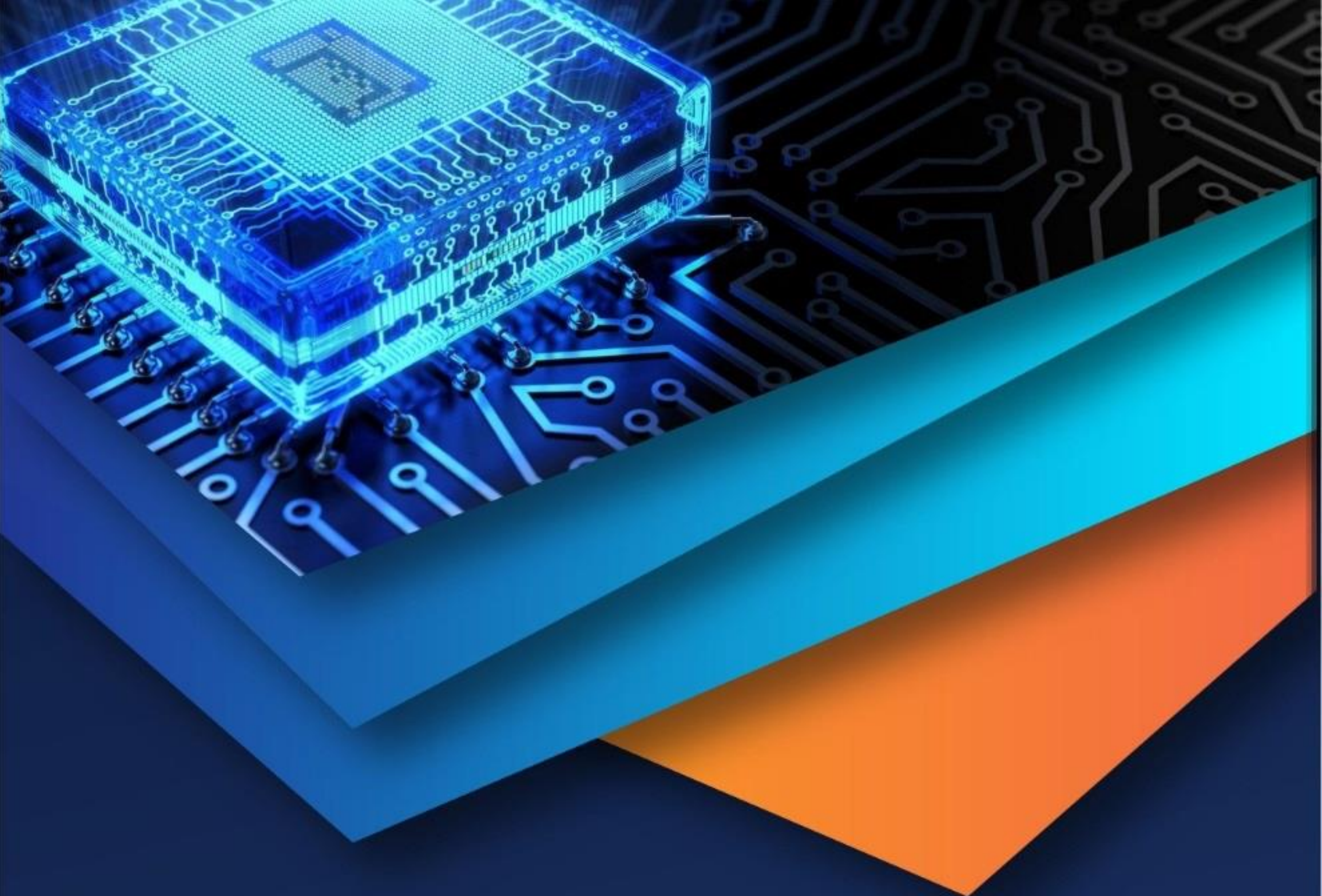
[21] Wang, J., Li, X., & Zhou, Y. (2025). Vision-based drone detection using temporal consistency modeling. *IEEE Transactions on Aerospace and Electronic Systems*. Advance online publication. https://doi.org/10.1109/TAES.2025.3456219

[22] He, Q., Sun, J., & Wang, P. (2024). Small UAV detection in cluttered backgrounds using attention-based deep learning. *Applied Intelligence, 54*(1), 203–219. https://doi.org/10.1007/s10489-023-05021-4

[23] Luo, Y., Chen, X., & Fan, G. (2025). Lightweight deep learning models for real-time UAV detection on edge devices. *Future Generation Computer Systems, 152*, 150–162. https://doi.org/10.1016/j.future.2024.01.015

[24] Islam, M. R., Rahman, M. S., & Hossain, M. A. (2024). Vision-based drone detection and tracking in complex environments. *Multimedia Tools and Applications, 83*, 25641–25663. https://doi.org/10.1007/s11042-024-18941-2

[25] Xu, L., Zhao, Y., & Li, D. (2025). Deep learning-driven UAV surveillance: Challenges, trends, and future directions. *Artificial Intelligence Review, 58*(3), 123. https://doi.org/10.1007/s10462-024-10618-9



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY