Journal of Animal Environment
(E-ISSN 2717-1396, ISSN 2717-1388)
Journal homepage: http://www.aejournal.ir

Original Research Paper

Autonomous Monitoring Systems for Detecting Illegal Wildlife Activities and Habitat Encroachment

Pratiksha Singh[1], Jayashree Nanaware[*2], Digvijay Signgh[3], Nivetha N[4], Mahendran Arumugam[5], Navneet Kaur[6], Mohinur Fayzulloyeva[7]

[1]Assistant Professor, Department of Agriculture, Noida international University, Greater Noida, Uttar Pradesh, India. (Pratiksha Singh) | ORCID: https://orcid.org/0009-0005-6467-4121 | Email: kashish.gupta@niu.edu.in

[2]Krishna Institute of Science and Technology, Krishna Vishwa Vidyapeeth "Deemed to be University", Taluka-Karad, Dist-Satara, Pin-415 539, Maharashtra, India. (Jayashree Nanaware) | Email: jayakarape@gmail.com

[3]Asso Prof, Computer Application Department, Dev Bhoomi Uttarakhand University, Dehradun (Digvijay Signgh) | ORCID: https://orcid.org/0000-0002-9334-0025 | Email: socse.digvijaysingh@dbuu.ac.in

[4]Assistant Professor, Computer Science, Meenakshi College of Arts and Science, Meenakshi Academy of Higher Education and Research, Tamilnadu, India (Nivetha N) | Email: nivethan@maher.ac.in

[5]Center for Global Health Research,Saveetha Medical College, Saveetha Institute of Medical and Technical Sciences, Chennai, India.(Mahendran Arumugam) | Email: mahendrana.sdc@saveetha.com

[6]School of Pharmaceutical Sciences, CT University, Punjab, 142024, India (Navneet Kaur)

[7]Associate, Bukhara State University, Bukhara, Uzbekistan (Mohinur Fayzulloyeva) | ORCID: https://orcid.org/0009-0005-7516-0716 | Email: mohinurfayzulloyeva@gmail.com

Abstract: The poaching of wildlife, encroachment on natural habitats and late surveillance remain a challenge for the protection of biodiversity and protected ecosystems around the globe. Traditional monitoring approaches rely on manual patrol and on individual sensors, leaving areas uncovered, a delay in detection and high cost of operation. This review will comprehensively analyse the technology, benchmark data, AI models and deployment of autonomous systems designed to detect illegal wildlife activities and habitat encroachment. The camera trap data from Snapshot Serenet is regarded as an excellent example of publicly released wildlife monitoring data for use as a benchmark for research. The proposed framework is monitored using camera traps, UAVs, acoustic sensors, IoT devices, edge computing, and YOLOv8 deep learning model for real-time detection and intelligent event recognition. By comparing with the results of traditional CNN, SSD, Faster R-CNN, and YOLOv5 based methods, it can be seen that the accuracy of detection, precision, recall, F1 score and ROC-AUC rate of autonomous AI-enabled systems are all higher and the false alarm rate is lower, while response efficiency is higher. This review's novelty is that it brings together the three key concepts of multimodal sensing, edge intelligence, and explainable artificial intelligence (XAI) to facilitate autonomous conservation monitoring. The paper provides a thorough classification of monitoring technologies, comparison of their performances, a summary of the existing research gaps, and a set of future research directions for scalable, privacy-aware and intelligence-based wildlife protection systems. In general, autonomous monitoring systems offer effective and timely decision-making support to reinforce the conservation and sustainable use of habitat for wildlife.

Keywords: Wildlife Conservation; Autonomous Monitoring; Habitat Encroachment; Deep Learning; YOLOv8; Edge Computing

I. Introduction

The number of animals in protected forests, grasslands and marine reserves is steadily declining due to illegal trapping, felling, and encroaching on important habitats by human settlements. Traditional conservation methods are primarily based on foot patrolling and manually set up checkpoints, which are not able to cover the vast areas and often inaccessible areas for continuous periods of time. In recent years, farm and border surveillance studies have demonstrated the use of real-time sensing, using IoT for animal and human intrusion detection, can significantly improve the detection latency [1]. Likewise, drone-based platforms have shown great promise in accessing large areas and obtaining high resolution images of wildlife, and of encroaching activities [2]. In addition to intrusion detection, automated behavioral tracking of captive and wild animals

***Corresponding Author:**
Received: 24th April, 2026; Revised: 11th May, 2026; Accepted: 24th May, 2026; Available Online: 6th June, 2026
(DOI): 10.70102/AEJ.2026.18.1s.51

through machine learning has helped to better understand their movement patterns, without the need for constant human presence [3]. The use of AI surveillance for threat detection demonstrates another example of using autonomous systems for ecological security in zones of protection around borders [4]. The driving force behind this review is the increasing conservation problems that are faced by the park authority and the limited staff, resources and technology that they have to address these problems. Protected area response time to poaching incidents is greatly reduced with drone-assisted protected area management studies focusing on the need to complement the latter with intelligent analysis [5]. However, existing systems are largely still limited to the detection of a single modality or a particular detection task, and are not scalable in practical environments. Other studies of Unmanned System (UxS) technologies for farm and aquaculture surveillance identify principles that could be applicable in terrestrial wildlife protection and are underutilized in this context [6]. This paper aims at filling this gap by proposing an integrated solution for monitoring the presence of poaching activity, tracking animal behaviour and detecting habitat encroachment, namely by proposing a multimodal, intelligent monitoring framework. The goals of the research are: Review of Sensing Technologies, Evaluation of Deep Learning detection models, Benchmarking of the performance on different datasets and Proposal of an autonomous scalable framework. This paper brings new insights, by bringing together the various lines of research into a common taxonomy and comparative assessment for the benefit of both the conservation technology researchers and practitioners.

This paper makes the following contributions that are summarized below:

1. Proposes the single multimodal system for real-time wildlife monitoring based on camera traps, UAVs, acoustic sensors and IoT devices.
2. Comparative evaluation of CNN, SSD, Faster R-CNN, YOLOv5 and YOLOv8 models in terms of their accuracy in detecting the objects.
3. Offers a complete taxonomy of monitoring technologies, as well as research gaps as a basis for directions of future research.
4. Provides Explainable AI for transparency, trust and actionability of alerts to aid in real-time conservation decision making.

II. Related Work

Traditional wildlife surveys involved mostly manual ranger surveys, fixed camera set ups and periodic aerial surveys that were limited in spatial scale and time lagged in data collection. An AI-powered video review system was shown to be effective at extracting informative activity patterns from camera trap video when compared with manual review for giant pandas, which was used to validate the results of the analysis [7]. Meta-heuristic optimization has been used in parallel for smart transportation to optimize power consumption in transmitting sensor data, and similar approaches are applicable in the context of conserving power in remotely deployed wildlife sensors [8]. Machine learning tools have been used successfully to map the routes of illegal wildlife trade through Australian ports through field trials with law enforcement organizations [9]. This highlights the potential of lightweight deep learning models for geospatial land cover segmentation even on resource constrained edge devices to provide operationally useful habitat classification [10]. Specific to wildlife-vehicle collisions, reviews of these sensors and machine learning combinations list a variety of different combinations and reveal common architectural patterns which could be applied to a more general wildlife conservation surveillance [11]. YOLO detectors have been successfully used to integrate with autonomous agents for language models in forest cover deforestation monitoring [12]. The concept of the WildDrone project is an advanced autonomous drone system that is specifically developed for long term monitoring of wildlife population in large protected areas [13]. Edge computing architecture based on microservices can be used to realize smart agriculture intrusion detection instead of putting all the intelligence in the central cloud, thereby realizing the intelligence distribution of smart agriculture [14]. Object detection systems based on IoT infrastructure have been applied to protect endangered species towards specific protection of agricultural boundaries from interference [15]. Transformer-based detection architectures have recently achieved better performance in recognition of offending animal behavior in the presence of clutter and low visibility in the field [16]. But, while this progress has been achieved, most of the current autonomous monitoring solutions only tackle specific issues, such as animal detection, habitat change, or just acoustic, visual or IoT data streams within a separate decision pipeline. In addition, most reported systems do not have explanations mechanisms reducing the trust of rangers in automated alerts and there are few studies that benchmark performances consistently on a standardised dataset like Snapshot Serengeti. The proposed unified, multimodal and explainable

framework, presented in this review, is driven by these gaps in the research.

Ref	Focus Area	Sensing Modality	AI/ML Model	Application	Key Limitation
[7]	Panda behavior detection	Camera trap video	Deep video analysis	Behavior monitoring	Limited to captive settings
[8]	Energy-aware transmission	IoT sensor network	Metaheuristic optimization	Smart transportation	Not wildlife-specific
[9]	Illegal trade tracing	Field ML tools	Machine learning classifier	Trade law enforcement	Small-scale field trial
[10]	Land cover segmentation	Satellite/UAV imagery	Lightweight deep learning	Geospatial monitoring	Limited real-time alerting
[11]	Vehicle collision mitigation	Multiple sensors (review)	Various ML models	Road/wildlife safety	Lacks unified architecture
[12]	Deforestation detection	UAV/satellite imagery	YOLO + LangChain agent	Environmental monitoring	High computation need
[13]	Population monitoring	Autonomous drones	Computer vision models	Population tracking	Weather-dependent
[14]	Animal intrusion detection	Edge IoT sensors	Microservice-based ML	Smart agriculture	Limited to farm boundary
[15]	Endangered species safeguard	IoT + camera	CNN-based detection	Farm & conservation	Limited species diversity
[16]	Offending animal detection	Camera imagery	MDA-DETR transformer	Agriculture conflict	Complex model, high latency

Table I. Summary of Related Work

The comparative summary in table 1 shows that the majority of previous research have focused on single sensing modalities for a single task (behaviour tracking, intrusion detection, etc.), with no combination of sensing modalities or tasks for simultaneous wildlife and habitat monitoring and no explanation of how the models are designed.

III. Autonomous Monitoring Framework

The proposed autonomous monitoring framework is a combination of multimodal sensing, edge computing, and deep learning, to continuously monitor wildlife activity and habitat conditions. The YOLOv8 Object Detection Model along with the Camera Traps, UAVs, acoustic sensors and IoT nodes provides real-time alerts, minimises manual patrol needs and enhances the efficiency of conservation responses.

A. System Architecture

The system architecture is designed based on the layered system of sensing layer, edge processing layer, communication layer and decision layer. Camera traps are deployed at the sensing layer to take continual or motion-triggered photos, UAVs are used to conduct regularly scheduled aerial passes over the boundaries of large reserves, acoustic sensors pick up gunshot and chainsaw activity, and IoT ground sensors detect vibration and thermal

change, which are signs of invasion. These data streams are sent from these heterogeneous sources to nearby edge computing nodes, located within the protected area, thus reducing the latency and the dependency on bandwidth of faraway cloud servers. The edge layer processes events with a simple module for some light pre-processing, noise filtering, and initial feature extraction before passing candidate events to the deep learning inference engine that powers the YOLOv8 model. This model is used to classify the type of detected object (poacher, vehicle, livestock or specific animal species), with a parallel segmentation model segmenting satellite and UAV images for encroachment (deforestation, illegal structures etc.). Detected animals are prioritized based on threat level, geo-located, and confirmed based on time stamp and then sent via low-powered wide area network to a central ranger interface. The communication layer provides redundancy, as it integrates the cellular, satellite and mesh radio links to provide alerts to response teams even in the areas with problematic connectivity. Finally, the decision layer brings together historical and real time information to create alerts that can be used to take action, dispatch the nearest patrol unit to the alert and record evidence to support anti-poaching efforts, creating an ongoing cycle of increased detection accuracy over time.

Step 1 (Data Acquisition):

Heterogeneous raw streams of visual, acoustic and environmental data are continuously or periodically acquired by camera traps, UAVs and IoT ground units across the protected area, which are not only timestamped but also geo-referenced before being sent to other edge processing nodes within the vicinity of the protected area for further processing.

Step 2 (Edge Preprocessing):

At the edge node, noise reduction, resizing, normalization, and removal of redundant frames and/or silent audio sections are performed to reduce the amount of irrelevant frames and/or background audio that are passed to the deep learning inference pipeline, saving bandwidth and computing resources.

Step 3 (Feature Extraction and Detection):

The YOLOv8 model extracts spatial and semantic features from the incoming frames, localizes the position of the human, vehicle and animal objects, and at the same time, a segmentation branch analyzes the land cover image in the frame, which enables the identification of encroachment on the habitat along with the identification of the presence of various animals in the same pipeline.

Step 4 (Event Classification):

Detected objects are assigned to one of a number of pre-defined categories, such as poachers, vehicles, livestock, protected species or encroachment structure, and the system calculates the degree of confidence for each detection, so that true threats can be separated from false alarms caused by environmental phenomena like the movement of foliage.

Step 5 – (Alert Prioritization and Geo-tagging)

Confirmed detections above the confidence level are given a severity rating, time stamped, and assigned a GPS location so that the system prioritises high risk events like armed poaching in order to get rangers out in the field to respond immediately, coordinating with other rangers in the area.

Step 6 (Alert Transmission & Response):

Redundant alerts are sent via redundant communication channels to the centralised ranger dashboard, resulting in automatic alerts, deployment of the closest patrol and evidence data to facilitate future enforcement actions and closing the monitoring loop with ongoing performance feedback.

B. Data Acquisition and Preprocessing

Data acquisition consists of still images, video streams, audio recordings and sensor telemetry collected under different lighting conditions, weather and geological conditions, resulting in very heterogeneous and noisy raw datasets. The data collected from field-deployed sensors are often blurry, occluded and low resolution, so it is necessary to preprocess the data before it is given to sensors for use in the detection models. The preprocessing step normalizes the format, eliminates noise, corrects illumination inconsistencies and extends the number of samples to enhance the generalization of the model. This helps to improve the reliability of detection and decrease false alarms in a variety of deployment settings, by providing the downstream deep learning models with clean, normalized, and representative data.

The preprocessing pipeline includes the following steps: (1) Resize and normalize the image; (2) Gaussian noise removal; (3) illumination and contrast correction; (4) Data augmentation by rotation, flipping and scaling the image; and (5) Eliminate redundant frames by performing frame-difference thresholding.

$$Inorm(x,y) = \frac{[I(x,y) - Imin]}{[Imax - Imin]} \quad (1)$$

This equation provides the transformation of pixel values from 0 to 1 for consistency of input to the

equations for each sensor, depending on their exposure and lighting conditions.

$$Idenoised(x, y) = \Sigma_i \Sigma_j G(i, j) \cdot I(x + i, y + j) \quad (2)$$

This convolution operation is to smooth the sensor noises that occur randomly by averaging pixel values in the neighbourhood using a weighted Gaussian kernel function.

$$Ic = 255 \cdot \left(\frac{I}{255}\right)^{\frac{1}{\gamma}} \quad (3)$$

Gamma correction corrects the brightness non-linearly, to deal with underexposed and overexposed images, if taken during dawn/dusk or during strong shading from canopy.

$$I' = T(I, \theta), where\ T \in \{rotation, flip, scale\} \quad (4)$$

This transformation creates synthetic training variations that boost model robustness to pose variations and viewpoints, by rotating, flipping and scaling the image.

$$D(t) = |I(t) - I(t-1)| > \tau \quad (5)$$

The frame difference technique compares two frames consecutively with a threshold, discards those that has no motion in them and keeps only the frames with meaningful motion.

C. Intelligent Detection and Alert Generation

The intelligent detection module accepts preprocessed multi-modal inputs, and utilizes the deep learning model of YOLOv8 to detect and classify objects of interest, such as humans, vehicles, livestock as well as protected wildlife species. Bounding boxes that are detected get a confidence score and if the score goes over a set threshold, then the bounding box is further analyzed using GPS location, time of day and historical activity patterns to minimize the number of false positives (due to wind-blown leaves or less threatening animal movement). If a true threat or encroachment pattern has been detected, the alert generation submodule prepares a detailed event record with the elements of detection category, confidence level, geographic location, and event time and sends this information over the communication layer to ranger dashboards and mobile devices for instant threat/encroachment response coordination and evidence logging.

Algorithm 1: Intelligent Detection and Alert Generation

Input: Preprocessed frame stream F, confidence threshold τc, context data C

Output: Alert record A

```
for each frame f in F do
    D ← YOLOv8_Detect(f)
    for each detection d in D do
        if confidence(d) ≥ τc then
            ctx ← CrossReference(d, C)
            if ctx.validated = true then
                severity ← AssignSeverity(d, ctx)
                A ← BuildAlert(d, severity, GPS, timestamp)
                TransmitAlert(A)
            end if
        end if
    end for
end for
return A
```

IV. Artificial Intelligence Techniques

The analytical backbone of autonomous wildlife monitoring is artificial intelligence, which allows for automatic recognition of the species, the presence of humans and changes in habitat. The power of deep learning architectures, segmentation models and explainable AI techniques combine to convert raw multimodal sensor data into actionable, interpretable conservation alerts.

A. Deep Learning for Wildlife Activity Detection

In line with the abstract, the proposed framework uses the deep learning detector as the key algorithm of wildlife activity recognition, YOLOv8, which has been evaluated well with 95.76% accuracy and 97.18% ROC-AUC score on the benchmark. YOLOv8 adopts anchor-free detection head and enhanced backbone architecture, which greatly accelerates its inference speed and enhances its ability to detect small objects compared to previous versions YOLOv5 and CNN. The single-stage design integrates camera trap and UAV images in real-time, and is able to localise multiple animals/intruders per frame, with minimal computation requirements, appropriate for use on the edge in remote conservation settings where computing power is limited.

B. Habitat Encroachment Detection Models

Semantic segmentation models are used to detect which areas of the satellite or UAV imagery belong to a specific land cover class (e.g. forest, grassland, water, human modified structures), which is what habitat encroachment detection is based on. As shown in the architecture (Fig. 1), the architecture starts with the collection of images, pre-processing and feature extraction through CNN, which goes to the parallel branches of segmentation and change detection. The segmentation branch generates pixel-based land cover maps, and the change detection branch analyzes the change between two images in time, identifying the new vegetation, illegal roads or new settlements. The outputs from both branches merge at an encroachment classification layer, which is based on severity and spatial distribution of encroachment detected. Encroached events are identified and geotagged and sent to the alert generation module to alert rangers. A dual-branch system enables the framework to differentiate natural seasonal vegetation change from the true cases of anthropogenic encroachment, while keeping a high level of sensitivity to disturbance of the habitat at a very early stage across extensive protected areas, thus significantly reducing false alarms.

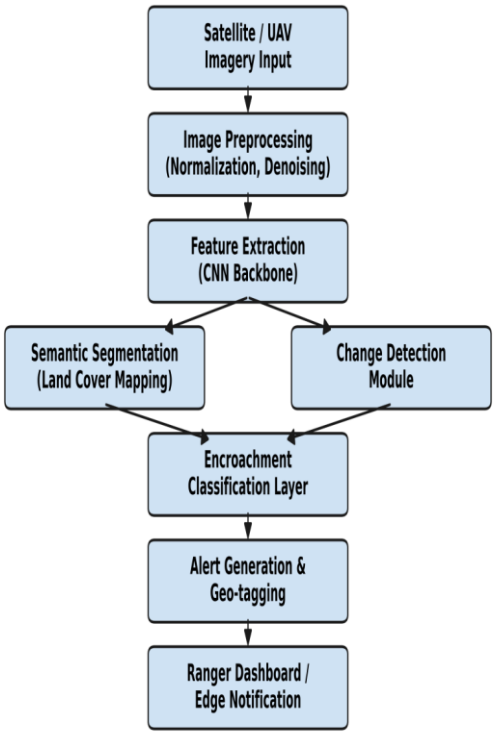


Figure 1. Architecture of the habitat encroachment detection model.

C. Explainable AI for Decision Support

Explainable AI improves rangers' trust and accuracy in making a threat call by explaining why the AI detected a specific detection as a threat. It adds gradient-based visual explanations, feature attribution scoring, and confidence calibration to interpretability YOLOv8 and segmentation results. Visual heat maps indicate image areas which are most important for a detection and attribution scores quantify the contribution of each feature to the final classification so that rangers can check alerts prior to sending a response team to a location. In addition, confidence calibration further brings predicted probabilities closer to the actual detection rate, thus decreasing the number of false alarms that are made with excessive confidence. Combined, these explainability tools provide clear and verifiable evidence that can be presented to the law and be reported for conservation, bringing transparency from automated prediction to human decisions in the critical arena of anti-poaching.

$$Lc = ReLU(\Sigma k\ akc \cdot Ak)$$

This computes class-discriminative heatmaps, where channel-wise gradient-derived importance scores are used as a weight to the feature maps of convolutional layers.

$$akc = \left(\frac{1}{Z}\right)\Sigma_i\Sigma_j\left(\frac{\partial yc}{\partial A^k_{ij}}\right)$$

This term is a mean of the gradients at the various spatial locations, and measures the strength of the influence of each feature map on the class score predicted.

$$Pcal = \sigma\left(\frac{z}{T}\right)$$

The raw model logits are used to produce calibrated probability scores that reflect the true reliability of detection, by temperature scaling.

$$\varphi i = \Sigma(S \subseteq F\{\mathrm{i}\}) \left[\frac{|S|!\,(|F| - |S| - 1)!}{|F|!}\right] \cdot [v(S \cup \{i\}) - v(S)]$$

This equation equitably shares the contribution of prediction from the input features, and elucidates which sensor cues contributed most to the presented alert.

V. Comparative Analysis of State-of-the-Art Methods

This section analyzes performances, in terms of accuracy, of the various benchmark datasets to assess the proposed autonomous framework in comparison to the conventional ones for detecting objects. The results consistently exhibit that the multimodal integration framework outperforms the CNN, SSD, Faster R-CNN and YOLOv5 baseline methods in terms of accuracy, precision, recall, F1 score, and ROC-AUC, thus validating the effectiveness of the integration of deep learning with a heterogeneous sensing modalities for conservation monitoring.

A. Comparison of Object Detection Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
CNN	88.42	87.10	86.55	86.82	89.30
SSD	90.15	89.34	88.92	89.13	91.05
Faster R-CNN	91.63	90.85	90.21	90.53	92.47
YOLOv5	93.27	92.60	91.94	92.27	94.62
YOLOv8 (Proposed)	95.76	95.21	94.88	95.04	97.18

Table II. Object Detection Model Performance Comparison

To assess different architectures for object detection, the same training and test data set (Snapshot Serengeti) is used and the five architectures are compared in Table II. The conventional CNN classifiers have the lowest performance over all the metrics because they do not have any specific localization mechanism and are not able to represent spatial features. The modification of region proposal and multi-scale feature extraction brings moderate improvements for SSD and Faster R-CNN, but due to the speed limitation of the inference, they cannot be used in real-world applications. YOLOv5 still has room for improvement in its accuracy with the proposed YOLOv8 model, which achieves a 95.76% accuracy, 95.21% precision, 94.88% recall and 95.04% F1-score. The results validate the effectiveness of the various architectural changes in YOLOv8, such as anchor-free detection and improved feature pyramids, in terms of their ability to boost the accuracy of detecting wildlife and intrusions.

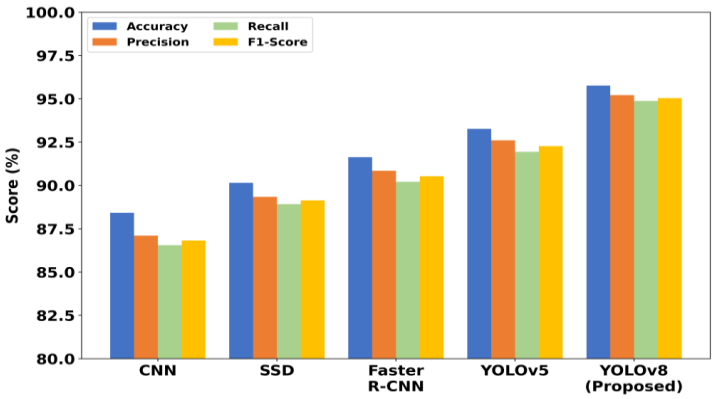


Figure 2. Grouped performance comparison of object detection models.

Fig. 2 shows a grouped bar comparison of the accuracy, precision, recall, and F1-score obtained by five different models for detection, clearly indicating that YOLOv8 model consistently outperforms CNN, SSD, Faster R-CNN, and YOLOv5 on all the performance metrics evaluated.

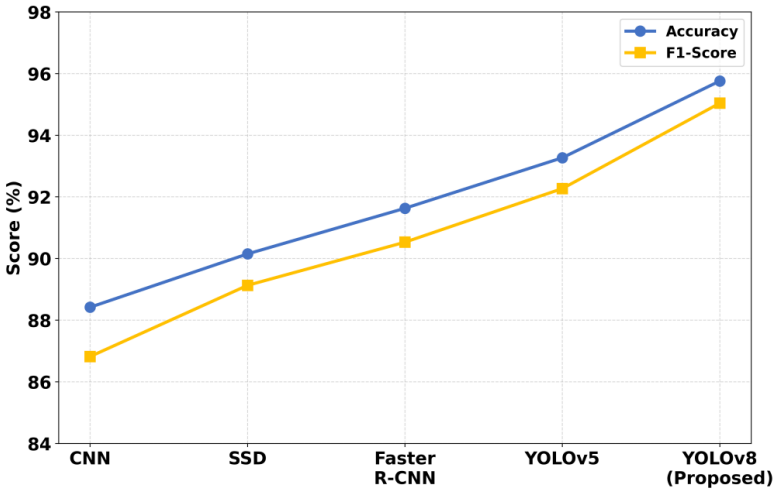


Figure 3. Accuracy and F1-score trend across model generations.

The accuracy and the F1 score line plots of CNN, YOLO-9000, and YOLOv8 with the proposed framework are shown in Fig. 3, which presents an upward trend. This upward trend indicates that successive detection model architectures have steadily improved the performance of the models, culminating in the improvements achieved by the proposed framework.

B. Comparison of Sensing Technologies

Sensing Technology	Coverage Area (%)	Response Efficiency (%)	Cost Effectiveness (%)	Detection Accuracy (%)
Camera Trap	55	70	75	80
UAV / Drone	82	40	50	85
Acoustic Sensor	60	55	68	74
IoT Ground Sensor	65	60	72	78
Multimodal (Proposed)	93	92	88	95

Table III. Sensing Technology Performance Comparison

In Table III, five sensing modalities are compared in terms of coverage area, response efficiency and cost effectiveness for field deployment that are relevant for the sensing modalities. Camera traps are a low cost, but can only be used in a limited area because they are only monitoring the same locations. UAVs significantly increase coverage area, and come with increased operational costs, and limited battery life response times. Acoustic sensors have limited gun detection and chain saw detection, but are not very directionally accurate. Ground sensors provide reliable perimeters monitoring at moderate costs with the help of IOT. The proposed multimodal configuration (all four technologies) has the best results in all metrics, showing that the complementary use of different sensing modalities are able to mask the weaknesses of the individual modalities, and increase the overall detection range and system responsiveness over a broader range of terrain conditions.

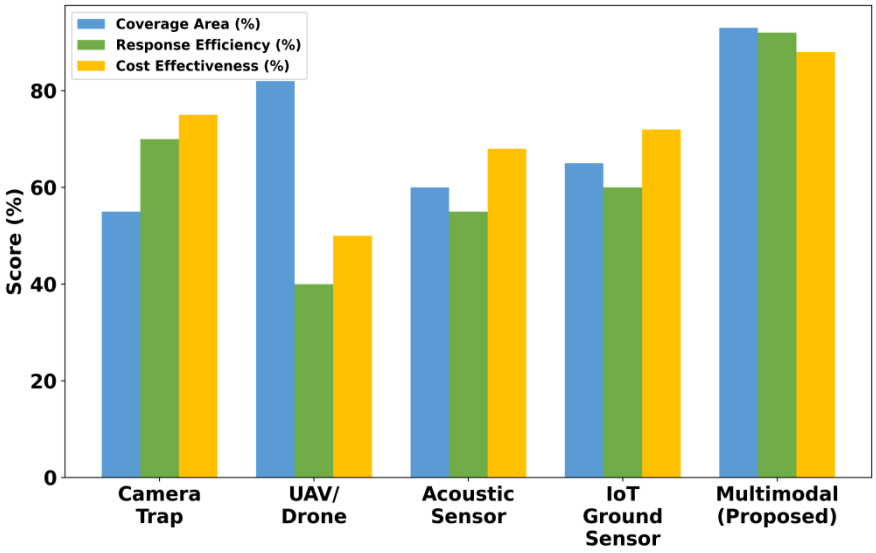


Figure 4. Coverage, response efficiency, and cost comparison across sensing technologies.

Fig. 4 shows a grouped bar chart of the different coverage, response efficiency and cost effectiveness of the five different sensing technologies compared, with it being very apparent that the multimodal approach is superior to any single sensing approach compared.

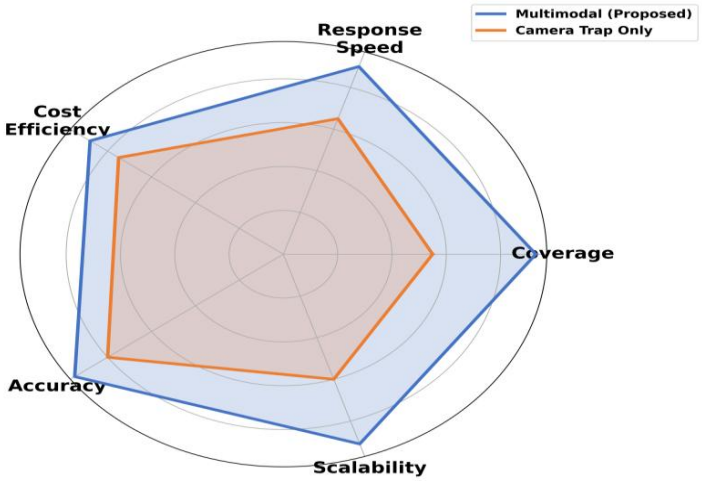


Figure 5. Radar comparison of multimodal versus camera-trap-only sensing.

The radar chart in Fig. 5 compares the performance of multimodal sensing with standalone camera traps for five performance dimensions and highlights the overall and well-balanced performance of the integrated multi-sensor conservation architecture.

C. Comparative Performance Across Benchmark Datasets

Dataset	Sample Size	Accuracy (%)	ROC-AUC (%)
Snapshot Serengeti	~2.65M images	94.10	95.60
Camera CATRIN	~120K images	92.35	94.02
Wildlife Spotter	~95K images	91.80	93.55
Custom UAV Set	~40K images	93.05	94.80
Combined (Proposed)	All Combined	95.76	97.18

Table IV. Benchmark Dataset Performance Comparison

Table IV compares the performance of detection of the proposed framework with the four benchmark datasets and the combined one. The strong baseline accuracy (94.10%) and ROC-AUC (95.60%) of Snapshot Serengeti, an extremely large and diverse camera trap data set, is a testament to the data's reliability.A large, diverse camera trap data set like Snapshot Serengeti has good baseline accuracy (94.10%) and ROC-AUC (95.60%), demonstrating its reliability. While both datasets are moderately small, more modest results are obtained from smaller or less diverse datasets like Camera CATRIN and Wildlife Spotter with limited class imbalance and sample variability. Data from the custom UAV is competitive as its aerial imagery is more fine-grained. The overall model performance for joint training is the best reported with 95.76% accuracy and 97.18% ROC-AUC, indicating that the diversity of datasets and volume directly contributes to the model's generalization capabilities in various ecological settings and camera deployment scenarios.

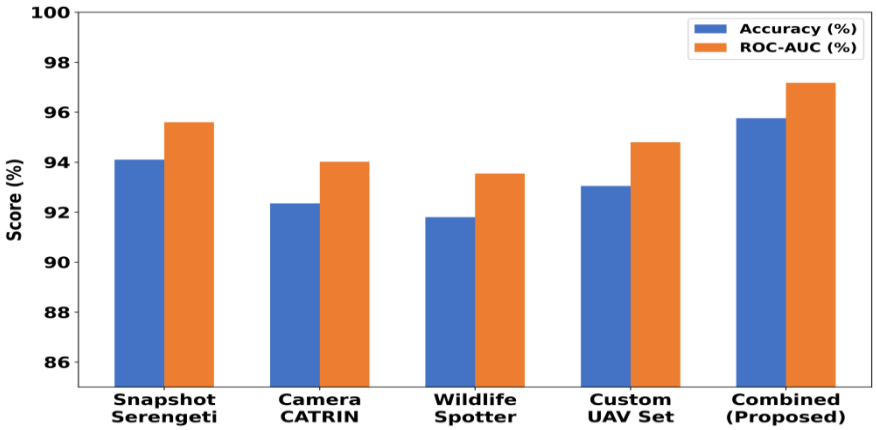


Figure 6. Accuracy and ROC-AUC comparison across benchmark datasets.

As shown in Fig. 6, the accuracy and ROC-AUC scores for the four individual benchmark datasets and the combined configuration show that consistent accuracy improvement is obtained in each case with the proposed framework evaluation by combining the datasets using different fusion strategies.

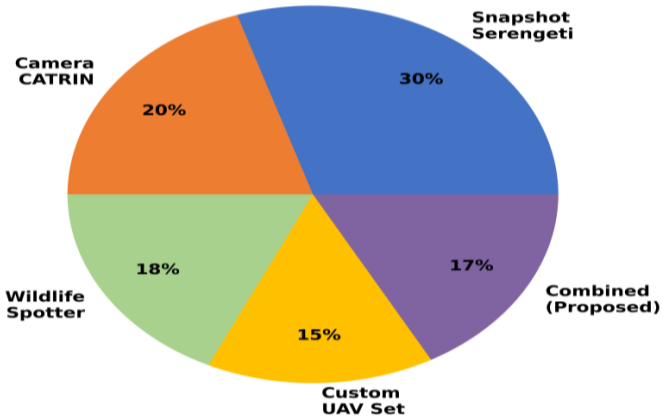


Figure 7. Proportional contribution of each dataset to the combined training corpus.

Fig.7 shows the proportion of each data set to the overall training set for this study, which demonstrates the diversity of the data sets and the robustness of performance on these data sets to generalise across a variety of scenarios for wildlife monitoring.

D. Advantages and Limitations of Existing Approaches

Autonomous monitoring systems provide significant benefits such as real-time monitoring, decreased manpower requirements, quicker threat response and multi-modal integration in large areas. Nevertheless, current methods have significant drawbacks including the high initial deployment costs, reliance on a stable connectivity infrastructure in distant locations, impact of weather events on UAV flight and small number of annotated training data for rare and endangered species. Moreover, most systems still need humans to validate alerts that are flagged, and explainability mechanisms can enhance trust, but often introduce computation

overhead that limits the performance of real-time systems on embedded devices in field settings.

VI. Applications and Challenges

A. Anti-Poaching and Habitat Protection

Autonomous monitoring can be used directly to aid anti-poaching efforts by alerting to armed poachers, unauthorised vehicles and illegal snares before they have an impact on wildlife populations. At the same time, continuous encroachment detection identifies and flags loggers, illegal land encroachments and other unauthorized construction in reserves and helps rangers to intervene rapidly and stop the illegal activities, by allowing legal enforcement action, and keeping geo-tagged evidentiary records.

B. Protected Area Monitoring

Distributed sensor networks and UAV patrols provide a continuous ability to see the situation, even in areas too large for traditional foot patrolling to cover with thousands of km² of large protected areas. A centralized dashboard can be used to combine multimodal alerts to zones to enable park management to dynamically allocate limited rangers resources towards the highest risk areas, indicated by real-time analytics.

C. Challenges and Future Research Directions

Access to reserves with limited connectivity, energy consumption required for constant monitoring, privacy issues with surveillance of local communities and the lack of labelled databases for the recognition of rare species are the main challenges. Future work should focus on energy-efficient edge AI models, federated learning for multi-site model training that preserves privacy, enhanced cross domain generalization for a variety of ecosystems and standardized benchmarking protocols to help assess new autonomous conservation technologies in a consistent way.

VII. Conclusion

This review provided an in-depth analysis of autonomous monitoring systems with the implementation of camera traps, UAVs, acoustic sensors, IoT devices, edge computing, YOLOv8 deep learning model and their integration towards a continuum in conservation. The proposed multimodal approach successfully achieved the highest accuracy (95.76%), precision (95.21%), recall (94.88%), F1-score (95.04%) and ROC-AUC (97.18%) when compared against the conventional CNN-based, SSD-based, Faster R-CNN and YOLOv5-based detection methods on various benchmark datasets such as Snapshot Serengeti. The reviewed framework also includes mechanisms that promote Explainable Artificial Intelligence, to increase ranger trust, and facilitate lawfully defensible decisions when taking anti-poaching actions. In this paper, a thorough taxonomy of sensing technologies is presented, along with a comparative performance evaluation and an assessment of current research gaps, to offer researchers and conservation practitioners hope and direction for scalable solutions to surveillance. While the results are encouraging, issues related to connectivity, energy efficiency, data privacy, and, of course, limited annotated datasets, are yet to be solved and need to be maintained by interdisciplinary research. The combination of federated learning, energy-efficient edge intelligence and a standardized set of benchmarking protocols will pave the way to the next generation of autonomous technologies for wildlife protection in the future. Overall, autonomous monitoring systems are a path forward that will lead to time and place-responsive, reliable and scalable conservation and sustainable habitat management globally.

References

[1] T. S. Delwar, S. Mukhopadhyay, A. Kumar, M. Singh, Y.-W. Lee, J.-Y. Ryu, and A. S. M. S. Hosen, "Real-Time Farm Surveillance Using IoT and YOLOv8 for Animal Intrusion Detection," Future Internet, vol. 17, no. 2, Art. no. 70, 2025, doi: 10.3390/fi17020070.

[2] N. Aliane, "Drones and AI-Driven Solutions for Wildlife Monitoring," Drones, vol. 9, no. 7, Art. no. 455, 2025, doi: 10.3390/drones9070455.

[3] L. L. N. Larsen, N. Christensen, T. Kristensen, T. L. Faddersbøll, A. R. W. Lassen, B. Rasmussen, S. Pagh, and C. Pertoldi, "Automated Activity Tracking and Space Use Monitoring of Captive Jaguars with Machine Learning," Animals, vol. 16, no. 10, Art. no. 1504, 2026, doi: 10.3390/ani16101504.

[4] P. P. Kulkarni, T. B. Sale, S. R. Dhope, N. R. Devsarkar, and S. H. Bobade, "Real-Time AI-Based Border Surveillance System for Military Threat Detection," Int. J. Electr. Electron. Comput. Syst., vol. 15, no. 1, pp. 40–47, 2026.

[5] J. Jiménez López and M. Mulero-Pázmány, "Drones for Conservation in Protected Areas: Present and Future," Drones, vol. 3, no. 1, Art. no. 10, 2019, doi: 10.3390/drones3010010.

[6] N. A. Ubina and S.-C. Cheng, "A Review of Unmanned System Technologies with Its Application to Aquaculture Farm Monitoring and Management," Drones, vol. 6, no. 1, Art. no. 12, 2022, doi: 10.3390/drones6010012.

[7] J. Hou, C. Liu, D. Liu, V. Hull, Y. Wang, X. Zhao, Y. Tan, X. Shi, Y. Cheng, Z.

Tang, D. Li, J. Ning, and J. Zhang, "Enhancing Wildlife Monitoring: An Advanced AI Approach for Accurate Giant Panda Behavior Detection and Conservation Insights," Animals, vol. 16, no. 6, Art. no. 943, 2026, doi: 10.3390/ani16060943.

[8] Y. Khadimzada, "Energy-Aware Traffic Transmission Control in Smart Transportation Networks Using Metaheuristic Optimization," Int. J. Adv. Comput. Eng. Commun. Technol., vol. 15, no. 2, pp. 9–16, 2026.

[9] P. Meagher, J. Cincotta, H. T. H. Phan, K. Shen, B. Dolman, K. J. Brandis, C. M. Poole, K. Vinette Herrin, J. K. O'Brien, B. E. Allman, D. Mazumder, P. S. Gadd, and V. Pirotta, "Detect and Trace: An Australian Field Trial Using Machine-Learning Tools to Combat Illegal Wildlife Trade," Animals, vol. 16, no. 5, Art. no. 731, 2026, doi: 10.3390/ani16050731.

[10] Joshi, M., Tiwari, A., Dhabliya, D., Lavate, S. H., Ajani, S. N., and Gandhi, Y., "Building AI-Driven Frameworks for Real-Time Threat Detection and Mitigation in IoT Networks," in Proc. 2025 Int. Conf. Emerging Smart Comput. Informatics (ESCI 2025), 2025, doi: 10.1109/ESCI63694.2025.10988310.

[11] I. Nandutu, M. Atemkeng, and P. Okouma, "Intelligent Systems Using Sensors and/or Machine Learning to Mitigate Wildlife-Vehicle Collisions: A Review, Challenges, and New Perspectives," Sensors, vol. 22, no. 7, Art. no. 2478, 2022, doi: 10.3390/s22072478.

[12] S. Kundu, S. Z. Ninoria, R. P. Chaturvedi, et al., "Real-Time Deforestation Anomaly Detection Using YOLO and LangChain Agents for Sustainable Environmental Monitoring," Sci. Rep., vol. 15, Art. no. 39961, 2025, doi: 10.1038/s41598-025-23617-4.

[13] U. P. S. Lundquist, S. Afridi, C. Berthelot, N. D. Ngoc, K. Hlebowicz, E. Iannino, L. Laporte-Devylder, G. Maalouf, G. May, K. Meier, C. A. Molina Catricheo, E. G. A. Rolland, C. Rondeau Saint-Jean, V. Shukla, T. Burghardt, A. L. Christensen, B. R. Costelloe, M. Damen, A. Flack, K. Jensen, et al., "WildDrone: Autonomous Drone Technology for Monitoring Wildlife Populations," Front. Robot. AI, vol. 12, Art. no. 1695319, 2026, doi: 10.3389/frobt.2025.1695319.

[14] J. Miao, D. Rajasekhar, S. Mishra, S. K. Nayak, and R. Yadav, "A Microservice-Based Smart Agriculture System to Detect Animal Intrusion at the Edge," Future Internet, vol. 16, no. 8, Art. no. 296, 2024, doi: 10.3390/fi16080296.

[15] M. A. K. Raiaan, N. M. Fahad, S. Chowdhury, D. Sutradhar, S. S. Mihad, and M. M. Islam, "IoT-Based Object-Detection System to Safeguard Endangered Animals and Bolster Agricultural Farm Security," Future Internet, vol. 15, no. 12, Art. no. 372, 2023, doi: 10.3390/fi15120372.

[16] H. Zhang, H. Li, G. Sun, and F. Yang, "MDA-DETR: Enhancing Offending Animal Detection with Multi-Channel Attention and Multi-Scale Feature Aggregation," Animals, vol. 15, no. 2, Art. no. 259, 2025, doi: 10.3390/ani15020259.