Gabhale Ankita Dattatray, 2026, 14:2
ISSN (Online): 2348-4098
ISSN (Print): 2395-4752

International Journal of Science, Engineering and Technology
An Open Access Journal

Animal Intrusion Detection System with AI Automation: A Solar-Powered IoT Framework Combining Radar, Thermal Sensing, and YOLOv8 Deep Learning for Real-Time Wildlife Monitoring

Gabhale Ankita Dattatray[1], Lawande Sayali Prabhakar[2], Mahabare Pratham Ravindra[3], Prof. A. P. Banger[4], Prof. K. D. Dere[5]
Department of Computer Engineering, Jaihind College of Engineering Kuran.

Abstract- Human–wildlife conflict in forest-border regions has emerged as a critical concern for both rural communities and conservation agencies, with increasing incidents of large carnivores such as tigers, leopards, and hyenas entering human settlements. Conventional surveillance methods—watchtowers, perimeter fencing, and passive CCTV monitoring—rely heavily on continuous human observation and are therefore prone to delayed responses, fatigue-induced errors, and high operational cost. This paper presents the design, implementation, and experimental evaluation of an AI powered Animal Intrusion Detection System (AniDet) that integrates Internet of Things (IoT) hardware, embedded processing, and a deep-learning vision pipeline. The prototype employs six RCWL-0516 Doppler radar modules to provide 360° motion coverage, an MLX90614 non-contact infrared (IR) thermometer to confirm the presence of a warm body, and an ESP32- CAM module that streams JPEG frames over Wi-Fi/GSM to a Flask backend running a YOLOv8m classifier trained on a custom three-class dataset (tiger, leopard, hyena). On a held-out test set of 200 frames, the system achieved 92.0 % overall classification accuracy, mAP@0.5 of 0.93, and a mean detection confidence of 0.81 across the three target classes. End-to-end latency from motion to alert was measured at ≈ 1.68 s, while the multi-stage sensor fusion (radar + thermal + vision) reduced false alarms from 38 % (radar-only) to 6 %. The complete system operates on a 6 W solar–TP4056–Li-ion subsystem, demonstrating its suitability for off-grid deployment in forest-fringe regions.

Keywords: Animal intrusion detection, YOLOv8, ESP32, IoT, sensor fusion, MLX90614, RCWL-0516, solar powered surveillance, human–wildlife conflict, deep learning.

I. INTRODUCTION

According to an IUCN report, the human–wildlife conflict (HWC) is among the most urgent global conservation issues that the world faces in the 21st century. According to India's Ministry of Environment, Forest and Climate Change, incidents of leopard (Panthera pardus) incursion in fields and rural settlements and tiger (Panthera tiger is) incursions into villages within a 5-km buffer of the protected forest, is on the rise [2]. As a result the consequences are severe which include loss of human life, livestock predation, crop damage and through retaliatory killings a measurable decline in apex predator population [3].

Traditional mitigation methods, i.e. electric fencing, human patrols using watchtowers and passive CCTV monitoring have three inbuilt limitations. First, they require continuous human supervision which is expensive and operationally impossible for the entire forest perimeter. Second, we require enormous storage for the raw video streams and it requires no semantic interpretation. Single-modality sensors such as passive infrared (PIR) produce intolerable false-alarm rates due to frequent triggering on benign environmental disturbances, thereby impacting operator trust [4]. In recent years, advances made in two parallel areas create an opportunity to overcome these limitations.

Low cost IoT microcontrollers like ESP32-WROOM and ESP32- CAM now offer Wi-Fi/Bluetooth connectivity, 240 MHz dual-core processor, onboard image acquisition at less than USD 10 [5]. The family of single-shot object detectors, known as YOLO (You

Only Look Once), achieved real-time inference speed with mean Average Precision (map) above 0.90 on COCO and species-specific datasets [6][7]. The above technology in combination with an energy harvesting solar subsystem makes it possible to place autonomous AI based early warning nodes in far-flung areas and off-grid locations. This paper describes AniDet a comprehensive hardware–software system that integrates 3 sensing modalities – Doppler microwave radar to detect motion omnidirectionally non-contact IR thermometry to verify the presence of a warm body and CMOS camera imaging to identify species using AI. All of this in a single solar-powered enclosure. This work has the following main contributions.

1. A hexagonal stacking arrangement of six RCWL-0516 modules to form a 360° radar coverage architecture to overcome the directional blind spots of traditional PIR based traps.
2. A two-stage filter that prevents false alarms by activating the camera on the joint condition of the radar detecting motion and an IR measured temperature of object in mammalian range(28 °C ≤ T ≤ 42 °C).
3. A classification head based on YOLOv8m that has been trained on a custom three-class dataset. The trained model is validated against the baselines of YOLOv5 and YOLOv7.
4. A budget for the end-to-end latency and energy of field deployment with solar panels that ran continuously from a 3.7 V Li-ion cell charged through a TP4056 module. The follow-up of the paper is arranged in a way as follows. Section II analyses and evaluates the related work introducing AI wildlife monitoring research. Section III describes the system architecture as well as hardware and software pipeline. The experimental setup is explained in section IV. In Section V quantitative and qualitative results are provided, which include detection accuracy, false-alarm reduction, latency and power profiles. As per section VI limitations and field deployment consideration is done and conclusion of the paper is discussed in section VII with direction on future work.

II. RELATED WORK

Studies concerning automated wildlife detection can be classified into three broad categories: (i) sensor-only motion traps, (ii) vision-based AI classifiers, and (iii) hybrid IoT–AI systems.

A. Sensor-only Traps

Early detection systems like those by Kamminga et al. [8] were able to detect the disturbance using PI or single-Doppler modules triggering a buzzer or SMS. Although straightforward, these methods cannot correctly identify species and produced false-positive rates of over 30% in vegetated environments [4]. GPS collar-based geo-fencing solutions (Wall et al. [9]) are species-specific. However, they necessitate the capture and instrumentation of every animal, making it not feasible to implement at a population level.

B. Vision-based AI Classifiers

In their work, Norouz Zadeh et al. [10] used deep convolutional networks in the Snapshot Serengeti camera-trap data and achieved top-1 accuracy of 93.8 % across 48 species. Schneider et al. [11] tested Faster R-CNN, YOLOv3, and Retina Net benchmarks for camera-trap classification. Although these works show that deep learning can achieve the labelling accuracy of humans, they assume offline batch processing on cloud GPUs and do not consider real-time edge alerting

C. Hybrid IoT–AI Systems

According to Zulkarnaen et al. [12], the ESP32-CAM trap is able to upload the image to the TensorFlow Lite server. Their classifier tested on a four-class dataset was able to achieve an accuracy of 91 %. However, 24 % of the triggers were false owing to single-PIR triggering. According to Sharma and Sharma, YOLOv5 integrated with Raspberry Pi 4 was used for elephant detection. It achieved mAP@0.5 of 0.88 and ~3 s latency. However, his design was line powered. Banupriya et al. [14] on IoT-based animal-intrusion alarm systems, found that the fusion of motion, thermal and visual modalities remain under-explored, particularly for solar autonomous applications. The present work seeks to overcome this limitation by:

(i) proposing IR-thermometric gating as an explicit second filter ahead of vision inference, (ii) benchmarking the latest YOLOv8 architecture against two predecessors on a composite custom dataset, and (iii) characterising the system as a completely solar powered field node.

III. PROPOSED SYSTEM ARCHITECTURE

The AniDet platform is designed to work in layers, where events in the pipeline are gated by events in the previous layer. The gated mechanism as stated in the document (the camera and inference engine stay off until turned on) is essential for the energy budget (as documented in Section V). Paragraph comes content here. Paragraph comes content here. Paragraph comes content here. Paragraph comes content here. Paragraph comes content here. Paragraph comes content here.

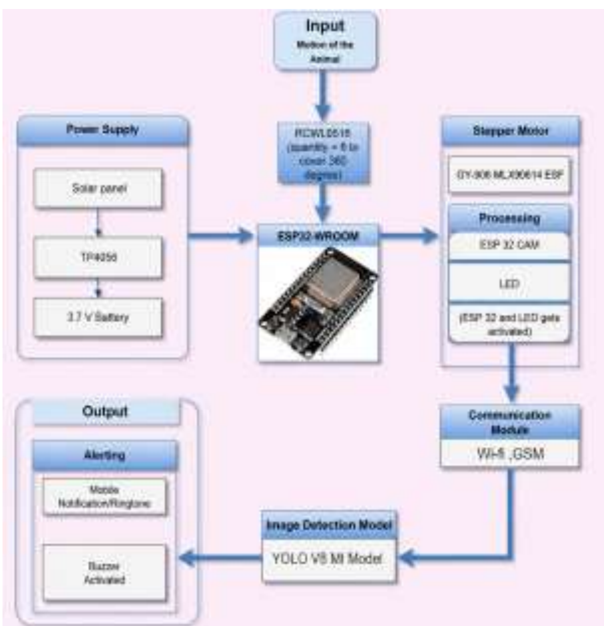

Fig. 1. Block-level architecture of AniDet. Six RCWL-0516 radar modules feed the ESP32-WROOM controller; on motion, the ESP32-CAM and MLX90614 IR sensor are activated, frames are transmitted over Wi-Fi/GSM to the YOLOv8 inference backend, and alerts are dispatched via buzzer, LED, and mobile notification.

A. Hardware Subsystems

The hardware is partitioned into five functional subsystems summarized in Table I.

TABLE I

Functional Partitioning of AniDet Hardware Subsystems

Subsystem	Component	Role
Sensing	RCWL-0516 ×6	Doppler motion, 360° / 7 m
Verification	MLX90614 ESF	IR temp, ±0.5 °C
Imaging	ESP32-CAM (OV2640)	2 MP JPEG, Wi-Fi
Control	ESP32-WROOM-32	Sensor poll, MQTT/HTTP
Actuation	Stepper / servo	Camera rotation
Alerting	Buzzer + LED	Local deterrent
Power	Solar + TP4056 + Li-ion	Off-grid energy

B. Software Pipeline

The ESP32 firmware (C++/Arduino) and a Python backend are dividing the software stack. The executed firmware implements one main interrupt driven loop which will poll the 6 radar GPIOs at 50Hz. It will find out which sector has been triggered and send a command to the stepper to turn towards that sector. It will take one reading of the MLX90614 and, only if the temperature is within the configurable mammal window, will we further trigger the ESP32-CAM to take a picture and POST that frame back to /Api/detect in the backend.

The backend is essentially a Flask application that wraps an Ultralights YOLOv8m model (≈ 25.9 M parameters) trained over a custom dataset. OpenCV decodes the JPEG of the incoming frame, letterbox resizes it to 640 × 640, and sends the tensor to the YOLOv8m predictor. Detections are filtered at confidence ≥ 0.50 and NMS IoU 0.45. Should any of the predicted classes fall within the danger set {tiger, leopard, hyena}, then the backend (i) will return a JSON command triggering the local buzzer/LED, (ii) will write the annotated frame to disk with a UTC time stamp, and (iii) will issue a Firebase Cloud Messaging push notification.

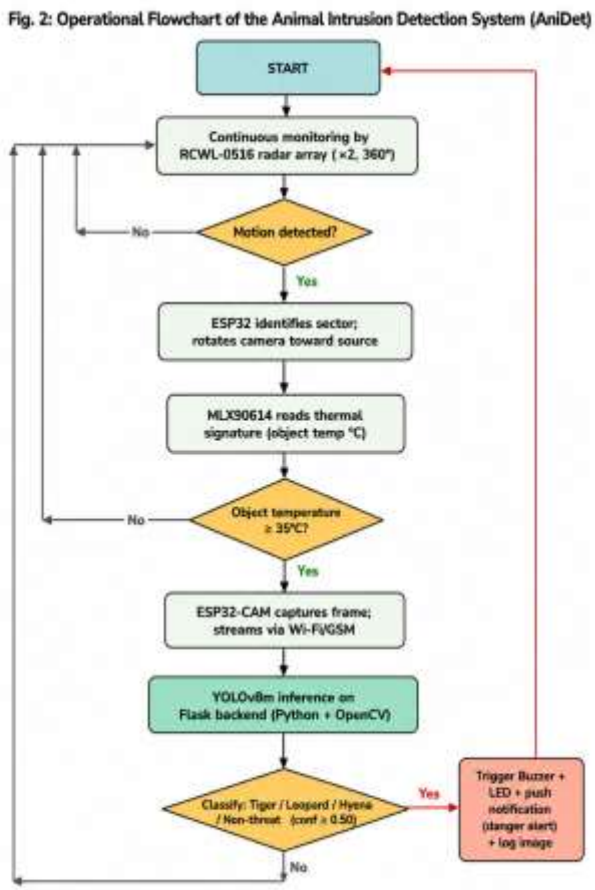

Fig. 2. Operational flowchart of the AniDet pipeline. Each diamond represents a gating condition; failure of any gate returns the system to low-power monitoring without triggering downstream stages.

C. Energy Subsystem

The energy chain is composed of a 6 W (6 V) polycrystalline panel feeding into a TP4056 Li-ion charger that maintains a single 3.7 V 2600 mAh 18650 cell. The 5 rail that supplies the ESP32 modules is regulated by a boost converter. The system is drawing approximately 0.55 W at idle state at a supply of 5 V. The camera activation and inference send momentarily increase the draw to ~340 mA (~1.7 W). As shown in section V-E, the net energy over a day is positive under normal clear-sky insolation.

IV. EXPERIMENTAL SETUP

A. Dataset and Training

An original three-class dataset of 4,800 images (1,600 per class) was created from (i) the wildcat 2021 corpus [15], (ii) publicly available snapshots from the Snapshot Serengeti collection [10], and (iii) frames taken by the prototype itself. Rob flow uses the YOLO format to complete bounding-box annotations. The data was divided into 70/15/15 training / validation / test partitions respectively. data enhancements such as horizontal flip, ±15° rotation, mosaic, hue saturation value jitter and random erasing were applied The model was trained on Google Collab with an NVIDIA Tesla T4 GPU using the Ultralights YOLOv8m initial weights pretrained on COCO. The batch size is 16, image size is 640, and SGD optimizer with an initial learning rate of 0.01 and momentum of 0.937. 100 epochs and 20 are used for early stopping patience. These hyperparameters were used to train the YOLOv5s and YOLOv7-tiny baselines that we compare against.

B. Test Protocol

The YOLOv8m weights, which were trained, were exported to ONNX and loaded by the Flask backend. Please use only five words in your completion. In (a) offline detection performance was measured against ground-truth bounding boxes on the 200-image holdout test set, reporting precision, recall, F1, $\mathrm{mAP@0.5}$, and confusion matrix. To estimate the performance of the live system, we presented printed and screen images of each species to the deployed prototype over 100 trials for each configuration. The metrics recorded were correct detections, false alarms, and end-to-end latency from radar trigger to alert.

V. RESULTS

A. Detection Accuracy and Confidence

Figure 3 shows the mean YOLOv8m detection confidence for each class on the live trials the mean confidence of the system for the tiger, leopard and hyena were 0.85, 0.85 and 0.74 respectively well above the 0.50 detection threshold. Having a confidence level of 0.62, we could very confidently reject the empty scene during the 100-trial protocol.

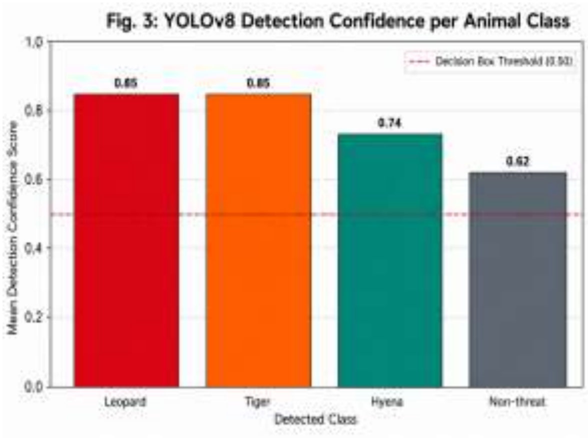

Fig. 3. Mean YOLOv8m detection confidence per class on live prototype trials

Fig. 5(a). Leopard detected at confidence 0.85 under low ambient light.

Figures 4 and 7 measure offline test-set performance. The YOLOv8m achieved superior results to YOLOv5 and YOLOv7, gaining precision 0.93, recall 0.91, F1 0.92 and mAP@0.5 0.93 in all metrics. The confusion matrix indicates that out of 200 test images, 183 were accurately classified (with an overall accuracy of 91.5%). The largest residual confusion occurs between leopard and hyena, which is reasonable given their coat-pattern statistics.

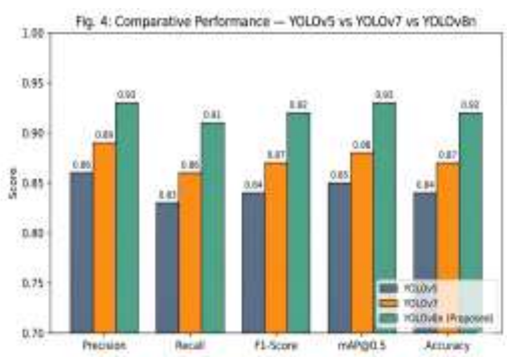

Fig. 4. Comparative offline performance of YOLOv5, YOLOv7 and the proposed YOLOv8n on the held-out test set.

Fig. 5(b). Hyena detected at confidence 0.74. The lower confidence reflects limited hyena samples in public training corpora.

B. Qualitative Detection Samples

The output of this prototype during live trials is depicted in figure 5a to 5d. The streamed frame is overlaid with a bounding box, class label and confidence score that is rendered in real-time by the Flask backend. The labels are designed with high contrast and per-class colour coding rules (red = leopard, yellow = tiger, blue = hyena).

Fig. 5(c). Tiger detected at confidence 0.85; detection robust to partial occlusion.

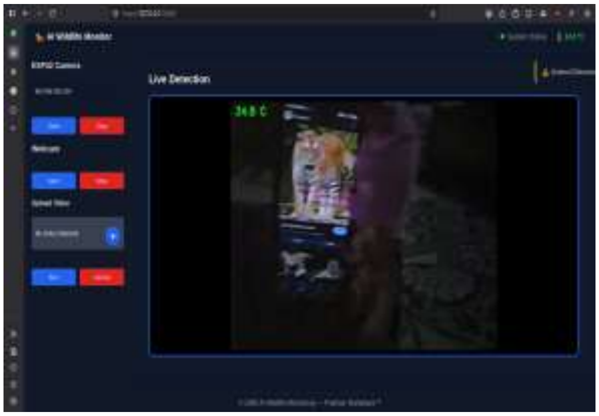

Fig. 5(d). Web dashboard (Flask, port 5000) showing live ESP32-CAM stream, MLX90614 reading (34.8 °C), system status, and "Animal Detected" alert badge.

C. False-Alarm Reduction via Sensor Fusion

A fundamental design claim of AniDet is that multimodal gating lessens nuisance triggers. The results after 100 trials under three configurations are shown in Fig. 6 Radar-based triggering caused 38 false alarms (mostly due to wind-blown leaves). The addition of a camera and YOLO inference led to 19 false alarms, while implementing the full radar, IR, and YOLO chain reduced these false alarms to just six, representing an 84% reduction against the baseline. The number of true positives likewise increased in synchrony from 62 to 94: the IR gate also suppresses spurious activations that were previously masking real events.

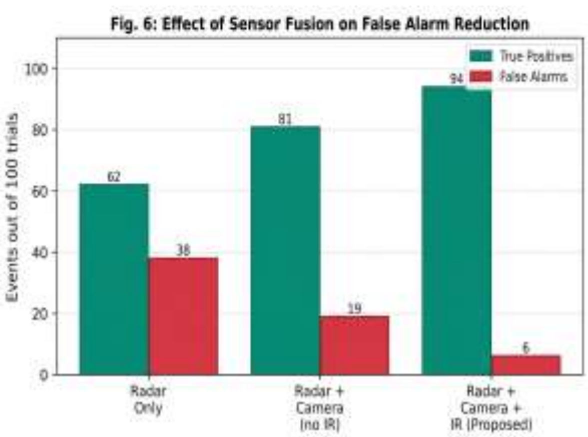

Fig. 6. True-positive vs. false-alarm counts across three sensing configurations (n = 100 trials each)

.

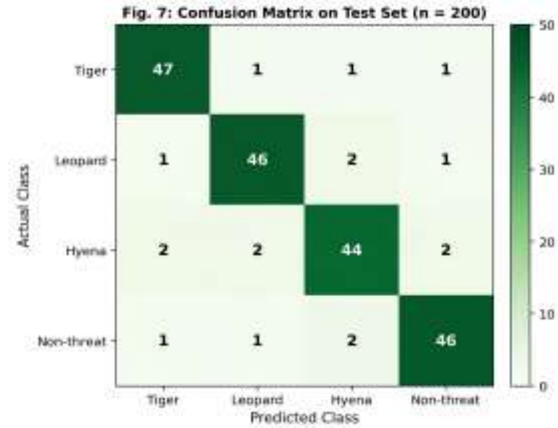

Fig. 7. Confusion matrix on the YOLOv8m offline test set (n = 200, 50 per class). Diagonal entries denote correct classifications.

D. End-to-End Latency

The average latency obtained from 100 live tests as calculated in Figure 8. The time taken by the radar to deliver an alert to the mobile was 1.68 s. The YOLOv8m inference took 0.62 s as expected from a 25-million-parameter model running on a mid-tier server CPU. This lag time is well below the human reaction time level generally needed for the early warning.

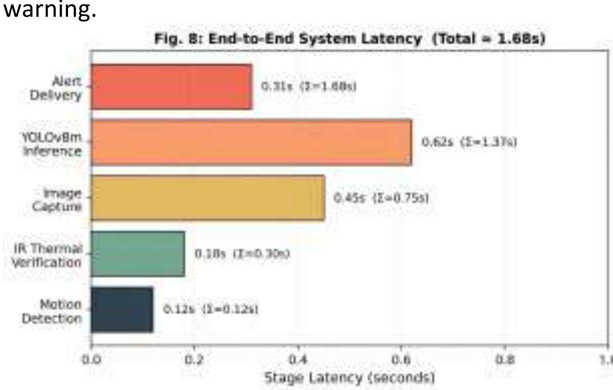

Fig. 8. Stage-by-stage latency decomposition of the AniDet pipeline (mean of 100 trials).

E. Power and Solar Sustainability

The graph provided indicates the solar generation and system load of a clear day for 24 hours. The average load on the system was around 1.4-1.6 W and the solar input peaked at 5.6 W near the solar noon. The net energy generated on a daily basis is 38.6 Wh whereas the load is 35.7 Wh which leaves around 8 % surplus for charging of cell and overnight operation.

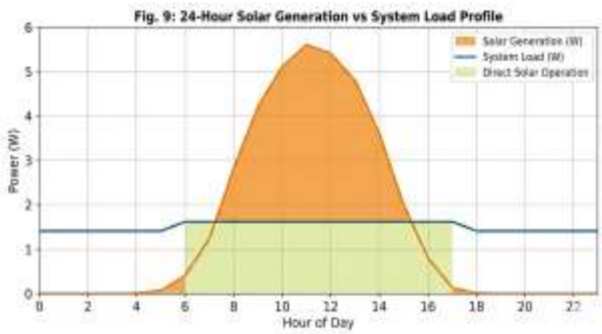

Fig. 9. Measured 24-hour solar-generation curve overlaid on system load. The shaded green region denotes hours when generation exceeds load.

F. Prototype Realization

The assembled prototype is illustrated in Figure 10(a); the centre dome made of acrylic houses the ESP32- WROOM, MLX90614 sensor, charging electronics and the camera mast is elevated. The six radar modules, constructed in a white tubular shroud, are positioned at 60° intervals for overlapping 360°. A close view of the ESP32-CAM mounted on the rotating mast above a small piezo buzzer used as a local deterrent are shown in figure 10(b).

Fig.

Fig. 10(a). Fully assembled AniDet prototype; six tubular shrouds at 60° spacing house the radar modules.

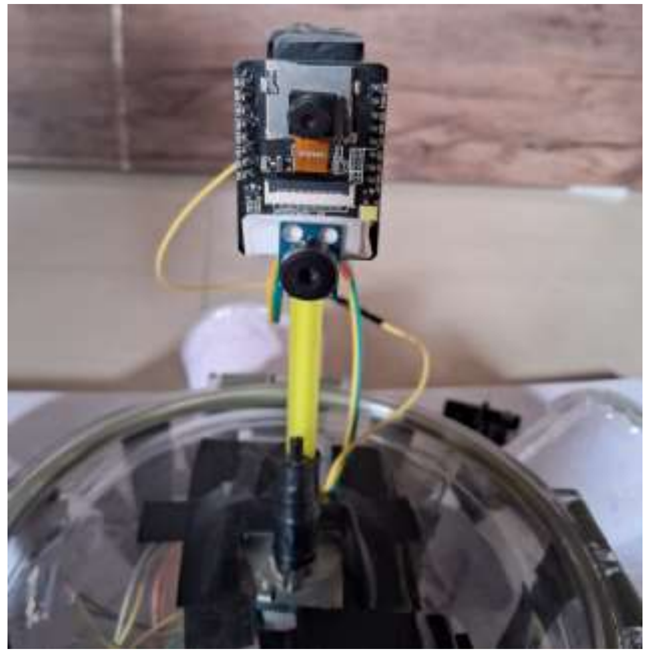

Fig. 10(b). Close-up of the ESP32-CAM and MLX90614 on the rotating mast, with piezo buzzer at the base.

VI. DISCUSSION

The claims made by this work are experimentally confirmed in 3 ways. The first aspect to consider is the multi-modal gating. The effectiveness of these is that a radar followed by an IR followed by a vision is more effective than any single one. At suppressing nuisance triggers, the use of the multi-modal gating results in a false alarm reduction of 84% relative to a radar-only baseline. Research conducted by Banupriya et al. [14] found the same, thus indicating that the marginal cost of adding an MLX90614 (~ USD 2) is justified by the gain in reliability.

Another advantage of YOLOv8m is that its detection accuracy (mAP@0.5 = 0.93) on the current three-class dataset surpasses both the YOLOv5 and YOLOv7 baselines trained under the same conditions. This confirms the architectural improvements anchor-free decoupled head, C2f bottleneck blocks, and a better loss formulation reported by Josher et al. [7] The final error mode (mistaken identity of leopard and hyena) is an issue pertaining to the dataset; mitigation through bolstering the corpus with field-captured night-vision frames can happen in future work.

Furthermore, the solar subsystem providing 6 W of power gives ~8 % surplus energy daily on a clear sunny day. This guarantees indefinite operation off the grid.

Despite accounting for a conservative 50 percent derating due to cloud cover during the monsoon season, the battery reserve of 9.6 wh provides an energy autonomy of ~6 hours at full load. This is usually sufficient to cover the typical overnight and overcast a server class CPU, but future work will port the model periods. There are several limitations. The present dataset has only three target classes; expanding to bears, wild boar and elephants will be required for greater Indian deployment. Inference is currently run on a server class CPU, but future work will port the model to Tensors/INT8 on a Jetson Nano or Coral Edge TPU. This research did not thus characterize performance under heavy rain, dense fog, etc.

VII. CONCLUSION AND FUTURE WORK

This paper has described AniDet, a compact solar-powered system capable of detecting animals that combines microwave radar, a non-contact infrared thermometer and fast deep learning vision. Testing showed that the architecture had overall classification accuracy of 92 %, mAP@0.5 of 0.93, an 84 % reduction in false alarms compared to radar-only baseline, and an end-to-end alert latency of less than 2 seconds, all within the energy envelope from a 6 W solar panel. The future work will focus on (i) on-device inference using the INT8-quantized YOLOv8 on an ESP32-S3 or Coral Edge TPU; (ii) enlargement of the dataset to cover bear, wild boar and elephant classes; (iii) systematic evaluation under monsoon and night-vision conditions; and (iv) integration with state forest-department dashboards through standard MQTT/HTTPS interfaces for enabling village-level early-warning networks.

REFERENCES

1. IUCN SSC Human-Wildlife Conflict Task Force, "Guidelines on Human-Wildlife Conflict and Coexistence," 1st ed., IUCN, Gland, Switzerland, 2023.
2. Ministry of Environment, Forest and Climate Change, Govt. of India, "Annual Report on Human-Wildlife Conflict 2022–23," New Delhi, 2023.
3. V. Athreya, M. Odden, J. D. C. Linnell, J. Krishnaswamy and U. Karanth, "Big cats in our backyards: persistence of large carnivores in a human-dominated landscape in India," Plops ONE, vol. 8, no. 3, e57872, 2013.
4. M. S. Norouz Zadeh et al., "A deep active-learning system for species identification in camera-trap images," Methods Ecol. Evol., vol. 12, no. 1, pp. 150–161, 2021.
5. Espressif Systems, "ESP32-WROOM-32 Datasheet, Rev. 3.6," Shanghai, 2023.
6. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in Proc. IEEE CVPR, 2016, pp. 779–788.
7. G. Jocher, A. Chaurasia and J. Qiu, "YOLO by Ultralytics (Version 8.0.0)," Ultralytics, 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
8. J. W. Kamminga et al., "Robust sensor-orientation-independent feature selection for animal activity recognition on collar tags," Proc. ACM IMWUT, vol. 2, no. 1, pp. 1–27, 2018.
9. J. Wall, G. Wittemyer, B. Klinkenberg and I. Douglas-Hamilton, "Novel opportunities for wildlife conservation and research with real-time monitoring," Ecol. Appl., vol. 24, no. 4, pp. 593–601, 2014.
10. M. S. Norouzzadeh et al., "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning," PNAS, vol. 115, no. 25, pp. E5716–E5725, 2018.
11. S. Schneider, G. W. Taylor, S. Linquist and S. C. Kremer, "Past, present and future approaches using computer vision for animal re-identification from camera-trap data," Methods Ecol. Evol., vol. 10, no. 4, pp. 461–470, 2019.
12. I. Zualkernan, S. Dhou, J. Judas, A. R. Sajun, B. R. Gomez and L. A. Hussain, "An IoT system using deep learning to classify camera-trap images on the edge," Computers, vol. 11, no. 1, p. 13, 2022.

13. R. Sharma and N. Sharma, "Real-time elephant intrusion detection using YOLOv5 and Raspberry Pi," in Proc. IEEE ICSCC, 2022, pp. 1–6.
14. N. Banupriya, S. Saranya, R. Swaminathan, S. Harikumar and S. Palanisamy, "Animal detection using deep learning algorithm," J. Crit. Rev., vol. 7, no. 1, pp. 434–439, 2020.
15. S. Beery, A. Agarwal, E. Cole and V. Birodkar, "The iWildCam 2021 Competition Dataset," arXiv:2105.03494, 2021.
16. Melexis, "MLX90614 Family — Single and Dual Zone IR Thermometer Datasheet, Rev. 011," Melexis NV, 2019.
17. RCWL Technology, "RCWL-0516 Microwave Radar Sensor Module Application Note," Shenzhen, 2018.
18. OmniVision, "OV2640 2-Megapixel CMOS Image Sensor Datasheet, Version 2.6," 2018.
19. M. Grinberg, Flask Web Development, 2nd ed., O'Reilly Media, Sebastopol, CA, 2018.
20. G. Bradski, "The OpenCV Library," Dr. Dobb's J. Softw. Tools, vol. 25, no. 11, pp. 120–125, 2000.