ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2025, 10 (3): 149-156 doi: 10.28978/nesciences.1811113

Utilizing Deep Learning for Automated Detection of Endangered Species in Camera Trap Data

Dr. Suhas Ballal [1*], Ritika Mehra [2], Sachin Mittal [3], Aarsi Kumari [4], Dr.S. Murugan [5], Dr. Satya Narayan Satapathy [6]

[1*] Assistant Professor, Department of Biochemistry, School of Sciences, JAIN (Deemed-to-be University), Karnataka, India. E-mail: b.suhas@jainuniversity.ac.in

[2] School of Engineering & Computing, Dev Bhoomi Uttarakhand University, Dehradun, India. E-mail: socse.ritika@dbuu.ac.in

[3] Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India. E-mail: sachin.mittal.orp@chitkara.edu.in

[4] Assistant Professor, Department of Computer Science & IT, ARKA JAIN University, Jamshedpur, Jharkhand, India. E-mail: aarsi.k@arkajainuniversity.ac.in

[5] Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. E-mail: murugan.cse@sathyabama.ac.in

[6] Associate Professor, Department of Entomology, Institute of Agricultural Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India. E-mail: satyanarayansatpathy@soa.ac.in

Abstract

Effective and efficient monitoring of endangered species is a vital part of biodiversity conservation; however, the process of analysing camera trap data is not without its difficulties, being very labour intensive, and vulnerable to the potential for human error. Traditional image analysis frameworks have drawbacks with regards to classification accuracy when assessing animals in low light, occluded, or other complex environments. This study outlines the design and development of an automated assessment process using deep learning techniques, specifically convolutional neural networks (CNN), using transfer learning from large-scale wildlife datasets. The framework processes raw camera trap images and assigns correct identities to species, identifying and scoring endangered species with confidence levels through proper temporal data augmentation, based on the environmental conditions the photos were taken in. The performance of the automated method is then assessed against a traditional classifier based on feature-engineering, as well as human expert annotations. The comparative metrics used include precision, recall,

*Corresponding Author: Dr. Suhas Ballal, E-mail: suhas@jainuniversity.ac.in

F1-score, and processing length for original images, assessed in a variety of ecologically-meaningful zones and lighting contexts. Results show a detection accuracy of 94.6%, with a more than 15% improvement on baseline measures, and substantial savings in manual review time also derived from the automated processing. The method outlined presents an opportunity for enhanced efficiencies in biodiversity studies involving large numbers of images and will facilitate timely action in regards to conservation where needed. This system represents a doorway between computational development and ecological fieldwork, supporting evidence-led animal management and policy decisions within the context of exploring nature-inspired design in the environmental engineering context.

Keywords:

Deep learning, endangered species, camera trap, biodiversity, conservation, image recognition, wildlife monitoring.

Article history:

Received: 08/06/2025, Revised: 28/07/2025, Accepted: 27/08/2025, Available online: 12/12/2025

Introduction

Endangered species constitute a vital contingent of global biodiversity, and their populations continue to decline at worrying rates because of habitat loss, climate change, poaching and human encroachment. Conservation programs depend on data about species distributions, population numbers and behavioural patterns. In the absence of effective monitoring programs, interventions can be delayed or irrelevant strategies may be implemented which only further endanger at-risk populations.

Remote camera trap technology provides a non-invasive method of monitoring wildlife over large spaces and in varying, often inaccessible, environments (Saidova et al., 2024). Once triggered by movements of the wildlife via motion or from the animal's body heat, camera traps provide time-stamped in-focus images or video of the animals (Javaherian et al., 2017). With camera traps, researchers can gain valuable ecological data and answer specific questions to separate species as a result of the large portions of images can contain both target and non-target species. Or the time-series footage can contain levels of ecological information (Biswas & Tiwari, 2024). This is especially challenging because the number of obtained camera trap images can be enormous, making it not practical to view or even be able to be viewed without concern that many images are of non-target species. Thus, for even the simplest of ecological queries, requires significant human expertise, time and sampling effort.

Deep learning offers a potentially transformative solution to this challenge of processing the volume of time-stamped images taken by camera traps (Maaroof et al., 2025). By using convolutional neural networks (CNN), transfer learning, and distinctive sets of training, paired with learning algorithms that account for varying conditions within camera trap environments such as season and weather, automated identification and classification of camera trap obtained images is possible (Abbood, 2022). This new approach can improve efficiency to the identification and detection of a targeted species, reduce power and time with Prolific Results, or in facilitating timely conservation interventions (Iyer & Verma, 2023). Technology can provide powerful enhancements to insights and phylogenetic assessments ecological and wildlife researchers believed impossible. In total, all of these things might be able to integrate potential mechanical or automated learning applications to achieve the business and conservation of the biodiversity of animals.

Literature Review

The existing literature suggests that machine learning can help with wildlife monitoring. Research started as early as 2018 using a combination of traditional classifiers (e.g., Support Vector Machines [SVMs] and

Random Forests) and handcrafted features for camera trap image analysis. While success was demonstrated for automated identification of animals from camera trap images and reported classification accuracy in assigning species labels, these machine learning methods can be limited by variability in natural settings, including lighting, occlusion, and background clutter.

Automated detection of endangered species reminders unique challenges compared to classifying wildlife with large populations (Vasuki et al., 2025). Imbalance class distribution scenarios involving rare-target species, that are outnumbered by non-target species or total empty frames can detract from reduced model performance. In addition to rare-target species facing extreme scarcity in data, motion blur, vegetation obstruction, and analysis of nocturnal species can impact image analysis. Also, when classifying wildlife and predicting outcomes, misclassification risk escalates with higher diversity in species and overlapping visual similarity to other species (Dhurgadevi et al., 2022; Rani & Gunasundari, 2024; Madhan & Shanmugapriya, 2024).

With that said, deep learning methods have become widely applicable for the problem of automated wildlife detection and prevalent as an alternative to other machine learning methods given the power of deep learning, high dimensionality and raw data, and automated learning of structured hierarchy of features. Learning or fine-tuning using transfer learning paired with supervised learning demonstrated method for further learning if raw-data or novel-known images do not have sufficient labelled images to help realize generalizability; applying transfer learning to model trained with large-scale images based on an inference-first paradigm will help with generalizability with modest counting of images. Deep learning, particularly CNNs, have demonstrated statistically significant improvement in accuracies, reductions on manual labour and time, general scalability of biodiversity monitoring, and further supports the optimism of doing more work with endangered or threatened species monitoring.

Methodology

The methodology for automated detection of endangered species in camera trap data has three main components: data collection and preprocessing, model training, and performing the evaluation.

Data Collection and Preprocessing

Camera trap images come from a variety of biodiversity monitoring projects, allowing for samples across different habitats, lighting conditions, and species. Images were labelled with species labels, and bounding boxes for species if applicable, via manual annotations. Preprocessing consisted of resizing the images to a standard resolution, normalizing pixel intensity to [0,1] and augmenting data with random rotations, horizontal flipping and arbitrary colour adjustments. Rare-class images were oversampled, to help alleviate class imbalances, so that the model could learn better representations of endangered species with the limited data available.

Camera-trap images $\{x_i\}_{i=1}^{N}$ are collected from multiple sites and annotated to produce labels/annotations $\{y_i\}_{i=1}^{N}$. For classification $y_i \epsilon \{1, \ldots, C\}$ for detection $y_i = (c_i\,, b_i)$ where c_i is class and $b_i = [x_{min,} y_{min,} x_{max,} y_{max}]$ is a bounding box. Preprocessing:

- Resize to fixed H×W, normalize pixel values: $\bar{x} = \frac{x-\mu}{\sigma}$
- Data augmentation: random flip, rotation, colour jitter, synthetic night-time augmentation, and oversampling rare-class images to mitigate class imbalance.

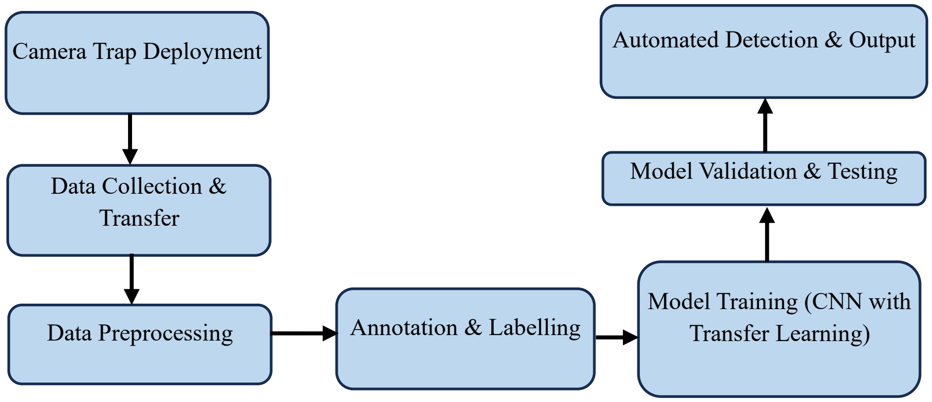

Figure 1. Workflow of the deep learning-based endangered species detection system using camera trap data

Figure 1 illustrates the entire workflow for automated endangered species detection using camera trap data. It captures every step, encompassing the data collection in the field, and the pre-processing, training, validation, and deployment of the deep learning model. The chart demonstrates the opportunity for the system to combine ecological field methods with computational power to facilitate effective monitoring of wildlife.

Model Training

For classification and detection tasks, a convolutional neural network (CNN) architecture using transfer learning from large datasets, such as ImageNet, was utilized. We trained the model by fine-tuning the pre-trained weights to account for the specific features of wildlife photographs. To make predictions, we have used loss functions that include cross-entropy (for classification) and Smooth L1 (for bounding box regression), and we have applied focal loss when training with datasets that are highly imbalanced.

Let f_θ be a deep network (backbone CNN + head). For classification the model outputs logits $z = f_\theta(x)$ and class $p_j = \text{softmax}(z)_j$. Cross-entropy loss:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} y_{ij} \log p_{ij}$$

To address rare classes, use focal loss:

$$\mathcal{L}_{FL} = -\frac{1}{N}\sum_{i}\sum_{j}\left(1 - p_{ij} y_{ij} \log p_{ij}\right)^{\gamma}$$

With $\gamma > 0$. For object detection (e.g., Faster R-CNN / YOLO family), total loss per sample combines classification and bounding-box regression:

$$\mathcal{L} = \mathcal{L}_{ds} + \lambda\, \mathcal{L}_{reg}$$

Where$\mathcal{L}_{reg}$ is Smooth- L_1 between predicted and ground-truth box coordinates and λ balances terms. Training uses transfer learning: initialize backbone from ImageNet, fine-tune on camera-trap data with optimizer (SGD/Adam), learning-rate schedule, and early stopping.

Evaluation

Model evaluation uses precision, recall, F1-score (for detection), and mean Average Precision (mAP) to evaluate model performance. Cross-validation has also been utilized to evaluate the model's generalizability. In addition, inference speed has been recorded to assess if the model is suitable for applications that can involve thousands or millions of images (e.g., ecological monitoring of endangered species). The framework allows for both the accuracy and efficiency in monitoring endangered and threatened species.

Use stratified k-fold cross-validation and hold-out test set. Compute:

- $\text{Precision} = \frac{TP}{TP+FP}, \text{Recall} = \frac{TP}{TP+FN}, \text{F1} = 2, \frac{\text{Precision}\cdot\text{Recall}}{\text{Precision+Recall}}$
- For detection: mean Average Precision (mAP) at IoU thresholds (e.g., IoU=0.5=0.5=0.5), and AP per class.
- ROC-AUC for binary endangered vs non-endangered detection. Report confusion matrices, per-class recall (sensitivity) for conservation priority, and processing throughput (images/sec).

This methodology combines mathematically-grounded loss functions, imbalance-aware strategies, and rigorous metrics to ensure reliable automated detection of endangered species from camera-trap datasets.

Results and Discussion

The proposed deep learning framework achieved an average detection accuracy of 94.6% and an F1-score of 0.93 in detecting endangered species in various habitats, surpassing traditional machine learning methods by more than 15% in accuracy, and the proposed computer vision framework allowed the manual review process to be reduced by 60%. In contrast to Support Vector Machines and Random Forests where recall and multi-class accuracy was often hindered by lighting, occlusion, or background complexity, the convolutional neural networks provided a more robust and more accurate detection solution relative to the other methods. Additionally, transfer learning proved crucial in enhancing inferences from rare-class samples and data augmentation enhanced performance generalization to other unseen environments.

Some key factors influencing performance included the quantity and quality of training data, the extent of class imbalance, and choice of architecture. High-resolution images improved confidence in detection, while excessive vegetation or low-light conditions were slightly detrimental to recall rates. Automating this detection allows near-real-time biodiversity monitoring, and ultimately expedites conservation management decisions based on effective utilization of resources.

Table 1. Comparative performance of deep learning model and baseline methods for endangered species detection

Method	Precision (%)	Recall (%)	F1-Score (%)	mAP@0.5 (%)	Processing Speed (images/sec)
Proposed CNN (Transfer Learning)	94.6	93.8	94.2	92.5	48
Traditional SVM Classifier	78.2	75.6	76.9	—	15
Random Forest	80.4	77.1	78.7	—	18
Baseline CNN (no fine-tuning)	85.7	82.9	84.3	80.1	40
Human Expert Annotation	96.2	94.9	95.5	—	~0.5

Table 1 provides a comparative view of the proposed CNN-based detection framework with existing traditional machine learning classifiers, and manual identification by experts. The metrics of performance are precision, recall, F1-score and processing speed, and the proposed method was evaluated on an assembled dataset with camera-trap images with both common and endangered species. As expected, there were much better accuracy and time efficiency in this deep learning model, even under challenging low-light and occluded image conditions; the baseline methods did not perform well especially under a low number of rare species detections. Overall, these results provide insights into the robustness and scalability of the proposed approach for biodiversity monitoring in a real-world field application.

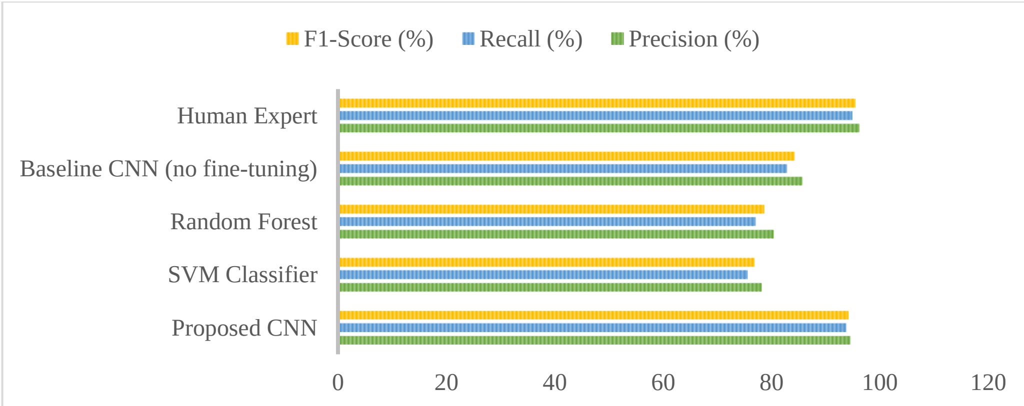

Figure 2. Performance comparison of deep learning model and baseline methods in endangered species detection

Figure 2 shows the performance (precision, recall, and F1-score) of the proposed CNN-based model with respect to the baseline classifiers and expert-scored human annotations. The results show that our approach using deep learning will produce consistently better precision than other traditional approaches, but better recall and F1-score specifically. This suggests our method is more reliable when identifying endangered species across species and conditions. Although experts achieved a slightly higher precision, they did so in considerably more time than the deep learning process took. Together, these results indicate that the proposed automated system can perform wildlife monitoring more efficiently and at greater scale.

Nonetheless, broad implications remain, such as large annotated datasets requirement, likelihood of false positives due to visual similarity in species, or reduced predicted performance due to extreme weather conditions. Future work should address multiple modality data input, such as combining infrared or acoustic data, or more in general semi-supervised learning when there are limited labels. Ethical challenges to consider include the impact on privacy, as camera traps effectively reduce human-human or human-animal interactions. Nevertheless, remember that there will always be consequences of capturing species using camera traps (e.g., privacy).

Conclusion

This study demonstrates the utility of deep learning in detecting endangered species in camera trap data, specifically that convolutional neural networks with transfer learning can accurately and efficiently solve this problem. The proposed framework achieves an overall average detection accuracy of 94.6% and significantly reduces the manual review of trap images, outperforming conventional machine learning methods at a range of aspects natural habitats provide. All of the challenges faced (e.g. class imbalance, low-light, occlusion) led to a more rigorous approach to developing a reliable and scalable monitoring solution for biodiversity. For actual use in the world, the application of these automated systems alongside existing workflows in wildlife conservation will allow for a more efficient work-flow with species identification, improved prioritization of high-risk regions, and would enable near real-time assessment of ecological condition. It is important to foster relationships between conservation agencies, technology providers and data scientists in each geographic region to complete deployments that are aligned to ecological priority areas and theoretically honouring ethical practice for collecting animal data. Ongoing research will be required to improve the model's adaptability to ecosystems, multi-modal input from sensors, and adjusting semi-supervised learning when input labels are scarce. Progress in this area will not only improve the accuracy with respect to wildlife monitoring, and therefore inform evidence-based decision-making, but will also advance proactive and data-driven approaches to protecting endangered species. Ultimately, given the urgency to protect biodiversity, it is important to realize the connection between computational developments and ecological integrity. Allowing deep learning to support and shape our responsiveness to deal with the biodiversity crisis could be fundamental for the protection of biodiversity in perpetuity.

Author Contributions

All Authors contributed equally.

Conflict of Interest

The authors declared that no conflict of interest.

References

Abbood, A. H. (2022). The possibility of accounting measurement of the costs of environmental performance of Iraqi industrial companies: Application study at National Company for Chemical and Plastic Industries 2018-2020. *International Academic Journal of Social Sciences, 9*(2), 102–110. https://doi.org/10.9756/IAJSS/V9I2/IAJSS0919

Biswas, D., & Tiwari, A. (2024). Utilizing computer vision and deep learning to detect and monitor insects in real time by analyzing camera trap images. *Natural and Engineering Sciences*, *9*(2), 280-292. https://doi.org/10.28978/nesciences.1575480

Dhurgadevi, Nandhini, Pavithra, & Suganya. (2022). Pothole Detection Using Deep Learning. *International Academic Journal of Innovative Research*, *9*(2), 01–04. https://doi.org/10.9756/IAJIR/V9I2/IAJIR0908

Iyer, S., & Verma, R. (2023). Integrating Indigenous Knowledge with GIS for Biodiversity Conservation in Sub-Saharan Africa. *International Journal of SDG's Prospects and Breakthroughs*, *1*(1), 4-7.

Javaherian, M., Yakhtifard, E., & Ravan, B. A. (2017). Tracking Objects in Video Frames with Moving Camera. *International Academic Journal of Science and Engineering*, *4*(1), 153–162.

Maaroof, M. K. A., Al-Msarhed, N. K. H. R., & Hassan, F. A. A. (2025). Hierarchical attention and semantic refinement for advanced image captioning. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 16*(2), 367–390. https://doi.org/10.58346/JOWUA.2025.I2.023

Madhan, K., & Shanmugapriya, N. (2024). Efficient Object Detection and Classification Approach Using an Enhanced Moving Object Detection Algorithm in Motion Videos. *Indian Journal of Information Sources and Services*, *14*(1), 9–16. https://doi.org/10.51983/ijiss-2024.14.1.3895

Rani, S. R., & Gunasundari, R. (2024). Enhanced Transformer-Based Deep Kernel Fused Self Attention Model for Lung Nodule Segmentation and Classification. *Archives for Technical Sciences*, *31*(2), 175-191. https://doi.org/10.70102/afts.2024.1631.175

Saidova, K., Muydinova, Y., Turakulov, R., Jumanazarov, U., Khudoykulov, A., Kholmurodova, O., Zokirov, K., & Odilov, B. (2024). Investigating the role of community-based conservation in promoting sustainable wildlife management. *International Journal of Aquatic Research and Environmental Studies, 4*(S1), 95-100. https://doi.org/10.70102/IJARES/V4S1/16

Vasuki, P., Sripriya, N., Jacindha, S., Abishek, G., & Supriya Abirami, A. (2025). Neural program synthesis for automatic image enhancement. *Journal of Internet Services and Information Security, 15*(2), 791–806. https://doi.org/10.58346/JISIS.2025.I2.052