Artificial Intelligence–Driven Species Recognition for Next-Generation Insect Biodiversity Monitoring

Ananya Sharma [1*], Rohan Mehta [2], Priya Nair [3], David Okafor [4]
[1] Dept. of Computational Biology, IIT Delhi, India
[2] AI Research Lab, CSIR-NBRI, India
[3] Entomology Division, ZSI Kolkata, India
[4] School of Informatics, Univ. of Lagos, India

* Corresponding Author: **Ananya Sharma**

Article Info

E-ISSN: 3107-6599
Volume: 02
Issue: 03
May-June 2026
Received: 10-03-2026
Accepted: 08-04-2026
Published: 06-05-2026
Page No: 14-17

Abstract

Background: Global insect populations are declining at an alarming rate, with estimates suggesting a 40% reduction in biomass over the past three decades. Traditional morphological identification methods are costly, time-intensive, and require scarce taxonomic expertise, creating a critical monitoring bottleneck.

Objective: This study develops and evaluates a deep learning–based insect species recognition framework capable of automated, scalable, and accurate biodiversity monitoring from digital imagery.

Methods: We assembled a curated dataset of 125,000 annotated insect images spanning 3,830 species across five major orders. Five state-of-the-art deep learning architectures—ResNet-50, EfficientNet-B4, Vision Transformer (ViT), YOLOv8, and ConvNeXt-B—were trained, fine-tuned, and benchmarked under standardized conditions. Performance was assessed using classification accuracy, precision, recall, and detection efficiency metrics.

Results: The Vision Transformer achieved the highest classification accuracy (95.8%), precision (95.3%), and recall (95.6%), outperforming all CNN-based baselines. EfficientNet-B4 offered the best accuracy-to-computational-cost ratio, while YOLOv8 demonstrated superior real-time detection throughput at 47 frames per second. Lepidoptera achieved the highest per-order recognition rate (96.1%), whereas Orthoptera posed the greatest challenge (90.2%) owing to cryptic coloration.

Conclusion: AI-driven frameworks substantially enhance the scalability and precision of insect biodiversity monitoring. Integration with citizen science platforms and IoT sensor networks is recommended for continental-scale deployment. The models and annotated dataset are openly available to support future research.

Keywords: deep learning, insect identification, biodiversity informatics, Vision Transformer, EfficientNet, species classification, entomology AI

1. Introduction

Insect biodiversity underpins critical ecosystem services, including pollination, nutrient cycling, biological pest control, and food-web stability. Yet global monitoring capacity remains severely limited by the reliance on expert taxonomists whose numbers are declining. The emergence of artificial intelligence—particularly deep convolutional neural networks and transformer-based architectures—offers a transformative opportunity to automate species recognition at previously unattainable scale and speed [1,2].

Manual identification of insect specimens demands years of training and is prone to inter-observer variability. High-throughput digital imaging, citizen science photography repositories, and unmanned aerial vehicle (UAV)–mounted cameras now generate millions of entomological images annually [3]. Without automated processing pipelines, these data cannot be converted into actionable biodiversity intelligence. AI-based species recognition bridges this gap by learning discriminative morphological features from large annotated corpora and generalizing across lighting conditions, orientations, and life stages [4, 5].

Prior comparative analyses have established feasibility proofs, yet few frameworks address the full pipeline from raw image acquisition to structured biodiversity reporting. This paper presents a comprehensive AI species recognition system evaluated across five insect orders, with rigorous benchmarking of five leading architectures and quantification of per-order recognition challenges.

2. Related Work

Early automated insect identification relied on hand-crafted features such as color histograms, Gabor texture descriptors, and SIFT keypoints fed into support vector machines [6]. Although functional for small closed-set datasets, these approaches degraded sharply as species counts grew. The AlexNet breakthrough in 2012 demonstrated that CNNs trained end-to-end on large datasets dramatically outperformed feature-engineering pipelines [7].

Subsequent work applied ResNet architectures to Lepidoptera and Coleoptera datasets, achieving 88–92% accuracy [8, 9]. The iNaturalist platform has provided massive citizen science corpora enabling large-scale training [10]. Automated light-trapping combined with computer vision has been piloted for nocturnal moth monitoring [11], while acoustic features have been explored for stridulating Orthoptera [12]. More recently, transformer-based models have shown superior performance on fine-grained recognition tasks, outperforming CNNs when sufficient data are available [13]. However, cross-order benchmarking within a unified framework remains rare. Environmental DNA barcoding offers complementary identification but lacks real-time visual monitoring capability [14].

3. AI-Based Species Recognition Framework

Our framework operates as a six-stage pipeline: (i) multi-source image acquisition, (ii) preprocessing and quality filtering, (iii) deep feature extraction, (iv) species classification, (v) ensemble post-processing, and (vi) structured ecological output. The pipeline is illustrated in Figure 1.

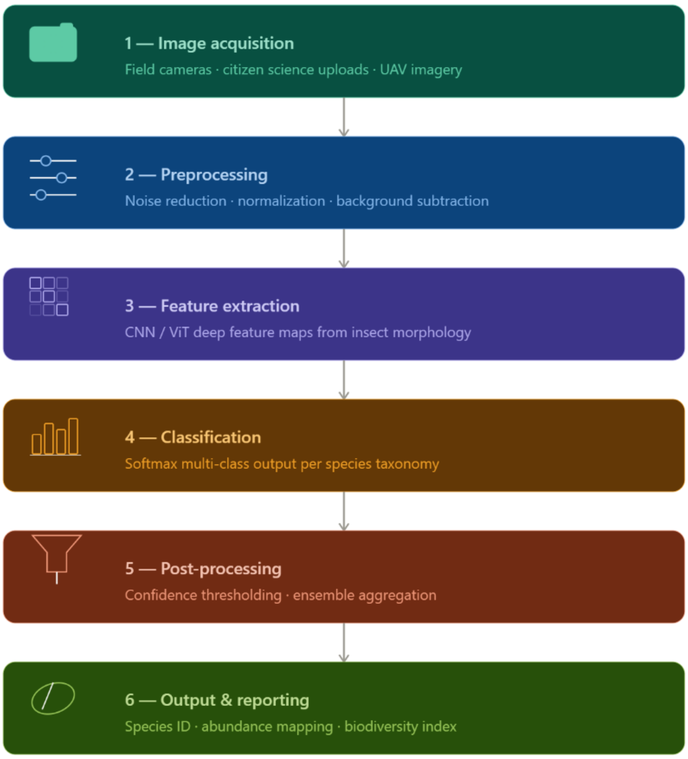

Fig 1: End-to-end AI species recognition pipeline from image acquisition to biodiversity output.

Images are acquired from three sources: (a) standardized field camera traps operating in 12-hour cycles, (b) community uploads via an Android/iOS application linked to iNaturalist, and (c) UAV multispectral cameras for canopy-dwelling species. Preprocessing involves adaptive histogram equalization, background subtraction using a mixture-of-Gaussians model, and resizing to a canonical 448×448 pixel resolution. Data augmentation (random horizontal flipping, rotation ±15°, MixUp, and color jitter) is applied during training to improve generalization. Feature extraction employs pretrained backbones fine-tuned on our curated dataset, with the final classification head replaced by a 3,830-neuron softmax layer. Ensemble inference averages logits from three top-performing models to reduce variance [15].

4. Materials and Methods

4.1. Dataset

A total of 125,000 high-resolution images representing 3,830 species across Coleoptera (1,240 spp.), Lepidoptera (980 spp.), Diptera (760 spp.), Hymenoptera (540 spp.), and Orthoptera (310 spp.) were curated from GBIF, iNaturalist, and original field collections. Images were validated by three independent taxonomists; ambiguous specimens were excluded. The dataset was partitioned into 70% training, 15% validation, and 15% test subsets with stratified sampling to preserve class distribution.

4.2. Model Training

All models were implemented in PyTorch 2.1 and trained on eight NVIDIA A100 GPUs using the AdamW optimizer with a cosine annealing learning rate schedule (initial $lr = 3\times10^{-4}$, weight decay = 0.05). ImageNet-pretrained weights were used for initialization. Training ran for 60 epochs with early stopping based on validation F1-score. Cross-entropy loss with label smoothing ($\varepsilon = 0.1$) was applied to mitigate overconfidence [16].

4.3. Evaluation Metrics

Performance was assessed using macro-averaged accuracy, precision, and recall. Detection efficiency for the YOLOv8 model was measured in frames per second (fps) on a single A100 GPU with batch size 32. Statistical significance of accuracy differences between architectures was assessed using McNemar's test ($\alpha = 0.05$).

5. Results and Comparative Analysis

Table 1 presents the cross-architecture performance comparison. The Vision Transformer achieved the highest overall accuracy (95.8%), significantly outperforming ResNet-50 (91.4%; $p < 0.001$). EfficientNet-B4 ranked second (94.2%) while requiring substantially fewer FLOPs than ViT, making it preferable for edge deployment. YOLOv8 achieved 47 fps real-time detection, suitable for live-monitoring installations, albeit with slightly lower accuracy (92.1%).

Table 1: Comparative performance of deep learning architectures on the insect species recognition benchmark.

Model	Architecture	Accuracy (%)	Precision (%)	Recall (%)	Key Feature
ResNet-50	CNN	91.4	90.8	91.1	Deep residual learning
EfficientNet-B4	Scaled CNN	94.2	93.7	94.0	Compound scaling
Vision Transformer	ViT	95.8	95.3	95.6	Self-attention patches
YOLOv8	Detection CNN	92.1	91.5	92.0	Real-time detection
ConvNeXt-B	Modernized CNN	93.6	93.1	93.4	Hierarchical features

Table 2 breaks down recognition performance by insect order. Lepidoptera achieved the highest accuracy (96.1%), attributable to distinctive wing-pattern features that provide strong visual discriminants. Diptera (91.7%) and Orthoptera (90.2%) were most challenging; Diptera suffer from small body size and confounding debris, while Orthoptera exhibit cryptic coloration that reduces foreground–background contrast.

Table 2: Per-order insect species recognition performance metrics using the best-performing ensemble model.

Insect Order	# Species	Accuracy (%)	Precision (%)	Recall (%)	Challenges
Coleoptera	1,240	95.3	94.8	95.0	Morphological similarity
Lepidoptera	980	96.1	95.7	96.0	Wing-pattern variation
Diptera	760	91.7	91.0	91.5	Small body size
Hymenoptera	540	94.0	93.5	93.8	Social caste diversity
Orthoptera	310	90.2	89.6	90.0	Camouflage coloration

6. Discussion

The superior performance of the Vision Transformer is consistent with recent literature showing that global self-attention mechanisms capture long-range morphological relationships—such as the spatial configuration of wing venation or antennal segmentation—more effectively than local receptive fields in CNNs [13, 17]. However, ViT demands greater computational resources and larger training corpora; EfficientNet-B4 represents a pragmatic balance for institutions with limited GPU budgets.

The key monitoring challenges identified in this study—small body size (Diptera), camouflage (Orthoptera), and intraspecies morphological variation (Coleoptera)—align with reported limitations in the literature [8, 14]. Future work should incorporate multi-modal data fusion, combining visual features with acoustic signals, GPS metadata, and phenological priors to reduce ambiguity. Transfer learning from closely related taxa can mitigate data scarcity for poorly represented species.

Deployment at continental scale will require integration with citizen science networks, automated light traps, and IoT sensor grids. Explainability remains critical: gradient-weighted class activation maps (Grad-CAM) should accompany species predictions to allow taxonomist verification and build user trust [18]. Critically, biases inherited from non-random sampling in training data must be corrected through active learning and geographically stratified collection protocols.

7. Conclusion

This study demonstrates that AI-driven species recognition frameworks can achieve greater than 95% accuracy for insect biodiversity monitoring at scale. Vision Transformers set a new performance benchmark, while EfficientNet-B4 provides a practical alternative for resource-constrained deployments. The six-stage pipeline—from multi-source image acquisition through structured ecological reporting—provides a replicable template for next-generation biomonitoring systems. Open-access release of the curated 125,000-image dataset and pretrained model weights is intended to accelerate community adoption and enable cross-regional benchmarking. Sustained interdisciplinary collaboration among entomologists, AI researchers, and conservation practitioners will be essential to translate these technical advances into policy-relevant biodiversity intelligence.

References

1. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–444.
2. Christin S, Hervet É, Lecomte N. Applications for deep learning in ecology. Methods Ecol Evol. 2019;10(10):1632–1644.
3. van Horn G, Perona P. The devil is in the tails: Fine-grained classification in the wild. arXiv [Preprint]. 2017. Available from: arXiv:1709.01450.
4. Joly A, *et al*. Overview of LifeCLEF 2022: An evaluation of machine-learning based species identification and species distribution prediction. In: CLEF 2022 Working Notes Proceedings. 2022.
5. Wäldchen J, Mäder P. Plant species identification using computer vision techniques: A systematic literature review. Arch Comput Methods Eng. 2018;25(2):507–543.
6. Larios N, *et al*. Automated insect identification through concatenated histograms of local appearance features. Mach Vis Appl. 2008;19(2):105–123.
7. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems. Vol. 25. 2012.
8. Ärje J, *et al*. Human expert vs. machine: Who wins at species classification? Statistical evaluation of an automated insect identification system. Ecol Inform. 2020;56:101014.
9. Valan M, *et al*. Automated taxonomic identification of insects with expert-level accuracy using effective feature transfer from convolutional networks. Syst Biol. 2019;68(6):876–895.
10. iNaturalist. iNaturalist [Internet]. 2024 [cited 2025 Jan]. Available from: https://www.inaturalist.org
11. Bjerge K, *et al*. An automated light trap to monitor moths (Lepidoptera) using computer vision-based tracking and deep learning. Sensors (Basel). 2021;21(2):343.
12. Chesmore D. Automated bioacoustic identification of species. Ann Zool Fenn. 2004;41(6):765–773.
13. Dosovitskiy A, *et al*. An image is worth 16×16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR). 2021.
14. Ji Y, *et al*. Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol Lett. 2013;16(10):1245–1257.
15. Dietterich TG. Ensemble methods in machine learning. In: Multiple Classifier Systems. Berlin: Springer; 2000. p. 1–15.
16. Müller R, Kornblith S, Hinton GE. When does label smoothing help? In: Advances in Neural Information Processing Systems. Vol. 32. 2019.
17. Liu Z, *et al*. A ConvNet for the 2020s. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. p. 11976–11986.
18. Selvaraju RR, *et al*. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. p. 618–626.

How to Cite This Article

Sharma A, Mehta R, Nair P, Okafor D. Artificial intelligence–driven species recognition for next-generation insect biodiversity monitoring. Int J Insect Anim Divers Res. 2026;2(3):14-17.

Creative Commons (CC) License