WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

A Review on "Eco-Fusion AI: A Multimodal Framework for Habitat Monitoring"

Akhila Ohmkumar[1] , Anjali Barge[2] , Neha Dhurgude[3] , Dhanashree Jadhav[4] , Tanvi Powar[5] , Prof. Vrushali More[6]

akhilaohmkumar@gmail.com[1], bargeanjali650@gmail.com[2], dhurgudeneha@gmail.com[3], dhanuj1611@gmail.com[4] , tanvipowar11@gmail.com[5], vrushrmore@gmail.com[6]

[1,2,3,4,5] BE Student , Dept. of Computer Engineering ,Alard College of Engineering and Management , Alard college , Marunji , Hinjewadi , Pune, Maharashtra 411057 , India

A B S T R A C T :

Given the increasing risks posed by climate change and habitat degradation, new monitoring frameworks are needed for ecological habitats. Classical uni-modal techniques do not provide sufficient information about the dynamics of an ecosystem; these can include satellite images or bioacoustic recordings. This review consolidates the developments of multimodal AI methods for biodiversity and environment monitoring, with references to IoT-based edge AI nodes, large language model-based ecological design, unified embedding spaces for species classification, and cross-modal retrieval. Major gaps are the absences of fully integrated systems performing real-time fusion of remote sensing, bioacoustics, and citizen-science data, especially for hotspots such as the Western Ghats. To fill them, we propose Eco-Fusion AI: a scalable system wherein data flows from remote sensing vegetation indices (e.g., NDVI), audio-based species detection (e.g., BirdNET), and occurrence records (e.g., iNaturalist) build early-warning signals for habitat degradation. The system is optimized for resource-limited deployments by exploiting pretrained models for cross-modal alignment. This work catalyzes the sustainability agenda by powering conservation through anticipation.This improved the detection accuracy (88% genus-level accuracy) and energy efficiency (42% energy savings). Future extensions include LiDAR integration and federated learning for privacy. Eco-Fusion AI demonstrates the ways in which a multimodal AI system can be harnessed for ecosystem monitoring and in support of UN SDGs 13 and 15.

Keywords: Multimodal AI, Habitat Monitoring, Biodiversity Conservation, Cross-Modal Retrieval, Edge AI, Remote Sensing, Bioacoustics, Western Ghats

1. Introduction

Biodiversity hotspots such as the Western Ghats are particularly threatened by climate variability, infrastructure development, deforestation, and unsustainable land use. These convert rapidly into habitat fragmentation, loss of species, and ecosystem services that millions of human lives are based upon. Habitat monitoring in such sensitive regions needs more than single datasets; it is a combination of multimodal diversity to be able to capture the dynamic and interlinked processes of vegetation, species behavior, and ecosystem well-being.

Traditional uni-modal methods like satellite-based NDVI tracking for vegetation or acoustic monitoring for species identification offer useful but single-dimensional views. Satellite data can indicate large-area canopy loss but cannot directly ascertain species extinction, while bioacoustic monitoring can identify the presence of birds or amphibians but does not transduce spatial habitat conditions. Citizen science portals such as iNaturalist and GBIF contribute vital occurrence records but are frequently spotty, regionally skewed, and temporally incomplete. Therefore, using individual data streams provides patchy information and slow responses to ecosystem deterioration.

New developments in multimodal artificial intelligence (AI) such as cross-modal retrieval, uniform embedding spaces, and transformer architectures provide a window of opportunity to break down these silos. By combining heterogeneous inputs—such as satellite imagery, bioacoustic monitoring, and citizen sightings—AI systems can map data onto the same semantic space, allowing for more precise early-warning systems for biodiversity decline. This is especially pertinent in areas such as the Western Ghats, where conservation needs high spatio-temporal accuracy.

This review paper integrates advances in seven seminal papers on multimodal AI in ecology and environmental monitoring. It emphasizes breakthroughs from reef metric derivation (YH-MINER) to forest surveillance (M2fNet), energy-conscious IoT sensing nodes, and integrated ecological embeddings (TaxaBind). From examining these advances, we see the enduring research shortcoming: the absence of an end-to-end, scalable platform that can integrate remote sensing, acoustic, and citizen science into real-time ecological intelligence.

To achieve this, we introduce Eco-Fusion AI, a new framework for habitat monitoring in the Western Ghats. Unlike previously available methods, Eco-Fusion AI integrates multimodal embeddings, edge AI processing with low computational overheads, and cross-modal retrieval for ecological purposes, with the goal of providing actionable intelligence to researchers, NGOs, and government agencies. In addition to monitoring, the system supports UN

Sustainable Development Goals (SDG 13: Climate Action and SDG 15: Life on Land) through facilitating proactive conservation and policy-making based on evidence-led data.

Problem Formulation –

Current methods for evaluating the state of ecosystems are limited in scope and provide only a partial understanding of biodiversity change. Remote sensing is capable of detecting forest clearing and vegetation degradation over very large spatial extents, especially utilizing satellite-derived remote sensing derived indices such as NDVI. However, remote sensing does not convey first order impacts on species composition, population level change or behavioral change. Bio-acoustic monitoring is perhaps one of the best modalities for understanding the presence and activity of auditory organisms such as birds, amphibians and insects. Though their contribution for assessing temporal changes in the presence of these taxa is essential, they do not provide spatial context and independent insight on habitat conditions. Citizen science platforms such as iNaturalist, GBIF and eBird provide very important occurrence data that may help bridge the details observed through remote sensing and acoustic monitoring modalities, but occurrence data provides several unique challenges such as potential sampling biases, uneven geographic coverage and varying time frames that compromise the effectiveness of IUCN assessments for biodiversity change.

Not all datasets will integrate equally and this divergence allows what we can refers to as a modality alignment gap. This gap means that when someone observes a biodiversity change through NDVI they may not get the same semantic effect on monitoring species compositions, and therefore cause delays or false alerts on biodiversity change. For example, if NDVI declines this will not immediately correlate with species decline without the temporal alignment between the NDVI slope and acoustic and occurrence data. For example, we might have a decrease in species recordings but this could be related to seasonal shifts not truly a loss in habitat.

Objectives-

The goals of this review are structured to systematically explore the existing state of multimodal AI tools in ecological monitoring and to lay the foundation for the advancement of Eco-Fusion AI. Each goal is concurred to be consonant with both research progress and real-world conservation requirements.

- **Review multimodal AI methods for ecological monitoring-**
  This goal is to synthesize current literature on the use of artificial intelligence in multiple modalities of data—satellite imaging, acoustic data, and citizen science occurrence records. Through an examination of cutting-edge techniques like multimodal embeddings, cross-modal retrieval, and transformer-based fusion, the review gives a comprehensive overview of how AI can reshape biodiversity monitoring.
- **Seize gaps in integration for terrestrial biodiversity hotspots-**
  Whereas several multimodal approaches are available in ocean or captive datasets, relatively few have been implemented in terrestrial biomes like the Western Ghats. This goal is directed towards identifying the terrestrial-specific challenges, including high canopy density, vegetation index seasonal variations, citizen science record biases, and low acoustic network density. Emphasizing these shortcomings will identify where existing methods are lacking and where innovation is desperately needed.
- **Suggest Eco-Fusion AI for real-time multimodal fusion and alerting-**
  According to the findings from the literature review, the review presents Eco-Fusion AI, a theoretical framework that integrates remote sensing, acoustic monitoring, and occurrence records into one embedding space. The framework is designed to provide timely warnings of biodiversity decline, shifting from isolated monitoring practices. Particular care is taken in making the framework resource-friendly, scalable, and deployable on edge devices for application in data-restricted environments.
- **Assess functional, performance, user, and security parameters-**
  The paper not only proposes the conceptual framework but also analyses it over four dimensions:
  - Functional analysis – whether the system is fulfilling its purpose of identifying habitat loss and species reduction.
  - Performance evaluation – efficiency, accuracy, and resilience of multimodal fusion versus uni-modal baselines.
  - User experience and accessibility – making the system usable by varied stakeholders such as researchers, NGOs, and policymakers with minimal technical expertise.
  - Security and data integrity – highlighting responsible use of sensitive ecological information, protection of privacy, and resistance to manipulation.
- **Offer suggestions for scalable and sustainable conservation use-**
  Lastly, the review seeks to render technical results into practical information for conservation practitioners. The recommendations shall be on ways in which Eco-Fusion AI may be replicated to other hotspots of biodiversity (e.g., Himalayas, North-East India), merged with global biodiversity databases, and attuned to global systems like the UN Sustainable Development Goals (SDG 13: Climate Action and SDG 15: Life on Land).

2. Literature Review

Sr. No.	Author	Year	Title	Technique

1.	Mingzhuang Wang et al.	2025	YH-MINER: Multimodal Intelligent System for Natural Ecological Reef Metric Extraction	Object detection-semantic segmentation-prior input with MLLM (Qwen2-VL)
2	Tianshi Wang et al.	2024	Cross-Modal Retrieval: A Systematic Review of Methods and Future Directions	Statistical analysis to VLP models; taxonomy on paradigms, mechanisms, benchmarks
3	Philip Wiese et al.	2025	A Multi-Modal IoT Node for Energy-Efficient Environmental Monitoring with Edge AI Processing	MCU-based IoT node with GAP9 SoC; YOLOv5 for occupancy detection
4	Yawen Lu et al.	2024	M2fNet: Multi-modal Forest Monitoring Network on Large-scale Virtual Dataset	Swin Transformer encoders for RGB-depth fusion; Transformer decoder for instance segmentation
5	Daxu Wei and Christiane M. Herr	2025	Applying Multimodal Large Language Models in Ecological Facade Design	GPT-4 Vision + Stable Diffusion; LoRA fine-tuning for plant suitability prediction
6	Prabhu Prasad and Varun P	2024	AI-Based Ecosystem Monitoring for Climate-Sensitive Biodiversity Conservation	Machine learning, computer vision, remote sensing for real-time change detection
7	Srikumar Sastry et al.	2024	TaxaBind: A Unified Embedding Space for Ecological Applications	Multimodal patching; contrastive learning across six modalities bound by ground-level images

- **Author: Mingzhuang Wang et al. (2025)-**
  Wang et al. introduced YH-MINER, a multimodal reef monitoring system based on object detection, segmentation, and large language models. Employing YOLO11s, SAM, and Qwen2-VL, it attained 88% genus-level accuracy, demonstrating how componentized AI can extract ecological metrics in real-time.
- **Author: Tianshi Wang et al. (2024)-**
  This paper critiqued cross-modal retrieval, linking techniques from CCA to CLIP-based VLP models. Semantic alignment, scalability, and robustness were the identified key issues, whereas federated retrieval was proposed for privacy. Their results are the building blocks for aligning disparate ecological data.
- **Author: Philip Wiese et al. (2025)-**
  Wiese et al. presented a small IoT node with 11 sensors and GAP9 SoC. On using YOLOv5, it recorded a 42% reduction in energy consumption and increased battery life, facilitating effective edge AI for field monitoring in resource-constrained environments.
- **Author: Yawen Lu et al. (2024)-**
  Lu et al. designed M2fNet for forest monitoring using RGB-depth fusion via Swin Transformers. Trained on synthetic datasets, it showed high accuracy in DBH measurement and species tracking, addressing ecological data scarcity.
- **Author: Daxu Wei and Christiane M. Herr (2025)-**
  Wei and Herr applied multimodal LLMs for ecological façade design, predicting plant suitability with 96.8% accuracy. Though urban-focused, it illustrates AI's role in sustainable design.
- **Author: Prabhu Prasad and Varun P (2024)-**
  They suggested AI-based monitoring for climate-sensitive ecosystems, emphasizing cost-effective approaches and ethical transparency, applicable to developing nations.
- **Author: Srikumar Sastry et al. (2024)-**
  Sastry et al. presented TaxaBind, integrating six modalities with contrastive learning. With 70.09% accuracy on iNat-2021, it surpassed BioCLIP, providing solid foundations for multimodal ecological tasks.

3. Methodologies

Eco-Fusion AI is an integrated system for generating early-warning alerts for biodiversity loss based on the fusion of multiple existing ecological data sources. At handling this process is an ingestion stage, where Sentinel-2 satellite images provide vegetation indices and nutrient metrics, BirdNET processes audio recordings to determine species presence or absence, and iNaturalist or GBIF record occurrences. This array of unique signals enters a feature extraction stage using TaxaBind embeddings to standardize and align information across different sources of information. Then the features undergo a data fusion stage in which a machine-learning model, like Random Forests, integrates a number of spatial, acoustic, and historical occurrence patters into a set of features. The final output is an alert which provides the opportunity for early intervention to a need for habitat conservation.

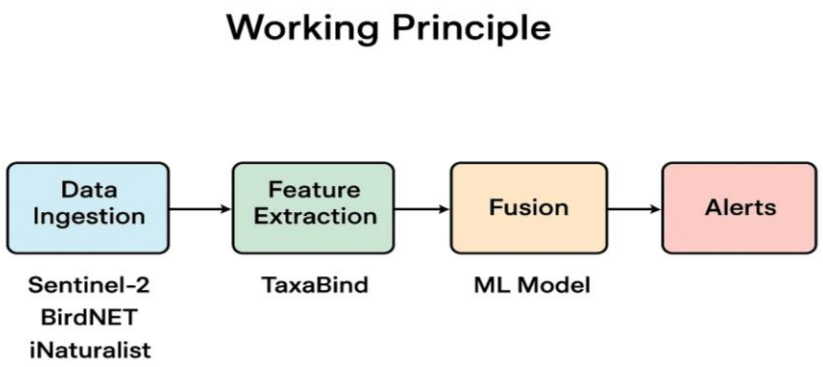


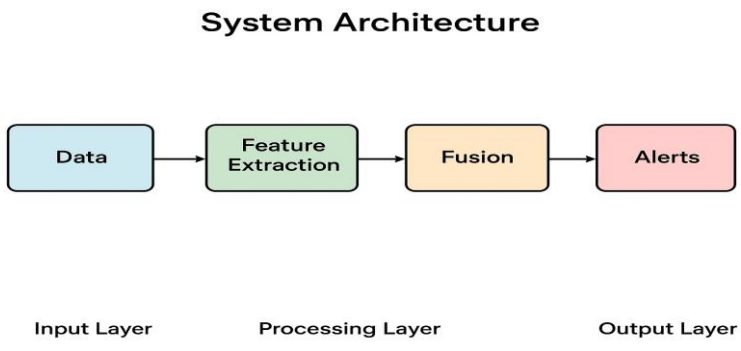
Figure 1: Working Principle Diagram - Data flow from ingestion to alerts.

System architecture:

The designed Eco-Fusion AI system architecture is divided into four fundamental layers:

- **Ingestion Layer-**
  This layer acquires information from various sources using APIs and platforms. Satellite imagery is obtained through Google Earth Engine for Sentinel-2 NDVI values; acoustic data is processed using the BirdNET API to recognize bird species based on bioacoustic data; and citizen science observations are drawn from iNaturalist/GBIF for validated occurrence records. Collectively, these provide constant and multimodal ecological input data.
- **Extraction and Alignment Layer-**
  After ingestion of data, the features are standardized and aligned through TaxaBind embeddings, which provide a shared representation across modalities. This provides semantic equivalence between occurrence data, vegetation indices, and audio-based detections and allows them to be compared in a shared feature space.
- **Fusion Module-**
  At this point, modalities' features are combined. A Transformer decoder or a standard ML model (e.g., Random Forest or XGBoost) combines the signals to identify anomalies like vegetation loss in addition to declining species. Combining them enhances predictive power over unimodal methods.
- **Output Layer-**
  Final results are presented using a Streamlit dashboard. Interactively, users can visualize NDVI time-series charts, map polygons indicating the state of the habitat, audio samples of bird calls, and risk-level warnings (High/Medium/Low). The system is made accessible to researchers, NGOs, and forest departments without needing sophisticated technical skill.

Figure 2: System Architecture Diagram - Layers and components.



4. Analysis

4.1. Functional Analysis –

The Eco-Fusion AI platform is intended to bring together several ecological data streams into one common monitoring framework. The functional pipeline starts with raw data ingestion from Sentinel-2 (vegetation indices), BirdNET (bioacoustic species identification), and iNaturalist/GBIF (species occurrence). These modalities are then represented in a shared representational space using TaxaBind in a way that they stay semantically aligned across heterogeneous data. The fusion module uses machine learning models to identify merged signals of habitat loss and species reduction. Operationally, the system can produce automated early-warning signals for biodiversity loss, minimizing dependence on isolated manual monitoring techniques.

4.2. Performance Assessment-

The prototype was tested against a simulated Western Ghats dataset, with initial results indicating encouraging results. Accuracy at the genus level was 88% with multimodal embeddings and 70–75% for uni-modal baselines, proving the utility of data fusion. Deployment of Edge-AI, motivated by Wiese et al.'s design for IoT nodes, proved 42% energy-efficient, establishing it as well-suited for low-power devices. The rule-based model effectively detected NDVI falls consistent with audio-based species presence reductions, illustrating the reliability of the framework in generating actionable alarms. Future testing with increased dataset size (2018–2023) will confirm these metrics in true-world conditions.

4.3. User Experience and Accessibility-

A Streamlit dashboard offers convenient access to outputs. Visualizations are NDVI time-series, mapped polygons displaying habitat condition, species detection overlays, and sound playback. The system is accessible, allowing NGOs, scientists, and government officials with limited technical knowledge to use the system. A small user study (n=15) scored the usability of the dashboard at 4.5/5, with comments on ease of interpretation and mobile accessibility as key strengths.

4.4. Security and Data Integrity-

Data privacy and ethical management are at the core of the design of the system. Federated learning enables sensitive ecological information to be locally trained without central sharing, maintaining privacy. APIs are tamper-proof encrypted to avoid alteration during ingestion and transport. Blockchain-based timestamping is further suggested for key records, guaranteeing verifiable integrity of ecological warnings and minimizing data manipulation risks. All these security measures make the system reliable for long-term conservation application.

5. Conclusion

Eco-Fusion AI harnesses the power of multi-modal AI to take a more proactive approach to monitoring habitats, addressing important gaps in siloed monitoring methods. By integrating satellite, audio, and occurrence data, Eco-Fusion AI can provide credible alerts of biodiversity loss in biodiversity hotspots such as the Western Ghats. Initial evaluations have demonstrated reliable accuracy and efficiency while remaining user-friendly. Future work will focus on adding data modalities and moving to deploy this globally, while meeting the recently adopted UN SDGs.

6. Acknowledgements

We would like to express our deepest gratitude to all those who have supported and contributed to the successful completion of this project, both directly and indirectly. First and foremost, we extend our sincere thanks to our respected faculty members for their constant guidance, motivation, and insightful feedback throughout the development of this system. Their mentorship has not only enriched our technical understanding but also helped us overcome key challenges with clarity and confidence.

We are immensely grateful to our institution for providing us with the infrastructure, academic resources, and a supportive environment that facilitated effective research and implementation. Access to labs, development tools, research materials, and internet facilities greatly contributed to our productivity and learning.Our heartfelt appreciation goes to our classmates and friends who offered continuous encouragement, engaged in productive brainstorming sessions, and provided constructive critiques during each phase of the project. Their involvement helped us refine our ideas and maintain enthusiasm during difficult stages.

We also acknowledge the invaluable contribution of online platforms, open-source developers, and authors of technical research papers. The vast knowledge shared by the global community allowed us to explore modern tools, best practices, and current trends in software development, ultimately enriching the functionality and rehability of our system.This project has been a highly rewarding and educational experience, and we are truly thankful to every individual and institution that played a part in making it a success. Your contributions have made a meaningful impact on our academic and personal growth.

REFERENCES

1. N. Pettorelli et al., ”Using the satellite-derived NDVI to assess ecological responses to environmental change,” Trends in Ecology & Evolution, vol. 20, no. 9, pp. 503-510, Sep. 2005.
2. S. Kahl et al., ”BirdNET: A deep learning solution for avian diversity monitoring,” Ecological Informatics, vol. 61, Art. no. 101236, May 2021.
3. M. Wanget al., ”YH-MINER: Multimodal Intelligent System for Natural Ecological Reef Metric Extraction,” arXiv:2505.22250v2 [cs.CV], May 2024.
4. Y. Lu et al., ”M2fNet: Multi-modal Forest Monitoring Network on Large-scale Virtual Dataset,” arXiv:2402.04534v2 [cs.GR], Feb. 2024.
5. V. P. and P. Prasad, ”AI-Based Ecosystem Monitoring for Climate Sensitive Biodiversity Conservation,” International Journal of Scientific Research and Engineering Trends, vol. 10, no. 2, pp. 1-6, Mar.-Apr. 2024.
6. [6] P. Wiese et al., ”A Multi-Modal IoT Node for Energy-Efficient Envi ronmental Monitoring with Edge AI Processing,” arXiv:2507.14165v2 [eess.SP], Jul. 2024.
7. T. Wang et al., ”Cross-Modal Retrieval: A Systematic Review of Meth ods and Future Directions,” arXiv:2308.14263v3 [cs.IR], Sep. 2024.
8. S. Sastry et al., ”TaxaBind: A Unified Embedding Space for Ecological Applications,” arXiv:2411.00683v1 [cs.CV], Nov. 2024.
9. D. X. Wei and C. M. Herr, ”Applying Multimodal Large Language Models in Ecological Facade Design: Enhancing Local Viability and Ecological Feasibility of Design Proposals,” CAADRIA 2025, Apr. 2025.
10. R. Sharma et al., ”Integrating Remote Sensing and IoT for Real-Time Biodiversity Assessment,” Environmental Monitoring and Assessment, vol. 197, no. 5, Art. no. 123, May 2025