A Multimodal Bayesian AI Framework for Multispecies Ecological Monitoring with Indigenous Knowledge Integration

Mahule Roy[1]
Subhas Roy[2]
[1]University of Oxford
[2]TATA Consumer Products Limited

Abstract

Contemporary conservation biology faces unprecedented challenges in monitoring biodiversity across scales while respecting cultural sovereignty and ecological complexity. We present an integrated AI framework that synergistically combines multimodal sensor networks with deeply embedded Indigenous Ecological Knowledge (IEK) to enable real-time, multispecies monitoring across heterogeneous ecosystems. Through rigorous evaluation across 12 benchmark datasets encompassing 3,800+ species and 4.2 million samples, our framework achieves state-of-the-art performance with 94.3% macro F1-score for species identification and 89.7% for behavioral classification. Critically, the Bayesian integration of IEK as structured priors improves rare species detection by 32.7% and enhances ecological plausibility by 41.2% compared to conventional data-driven approaches. Systematic ablation studies reveal the essential nature of multimodal fusion, with combined acoustic-visual-environmental models outperforming unimodal baselines by 23.8-45.6%. The architecture maintains practical deployability, with optimized edge models achieving 15.3 FPS while preserving 91.2% of cloud-level accuracy. This work establishes a new paradigm for ethically grounded, ecologically intelligent conservation technologies that bridge artificial and ancestral intelligence.

Introduction

The accelerating biodiversity crisis demands transformative ecological monitoring approaches (Dirzo et al., 2014). Current conservation AI faces fundamental limitations: single-species focus, inadequate ecological context handling, poor rare species performance, and limited cultural integration (Christin et al., 2019). These constraints are particularly acute in biodiverse regions with logistical barriers (Stevenson et al., 2023). Indigenous knowledge systems offer millennia of observational expertise to enhance AI ecological grounding (Fernández-Llamazares et al., 2016), yet current integration often treats Indigenous knowledge as data points rather than epistemological frameworks (Reo et al., 2017), representing both ethical and technical limitations.

Research Challenges and Contributions

Ecological AI systems face fundamental challenges in multispecies complexity, data scarcity, contextual awareness, operational deployment, and ethical integration. Our framework addresses these through hierarchical attention mechanisms and graph neural networks for multispecies monitoring, IEK-informed Bayesian priors and meta-learning for data-scarce scenarios, and temporal graph networks for ecological context. We overcome deployment barriers via neural architecture search and adaptive computation, while ensuring ethical integrity through co-design methodologies and federated learning. Our work makes four key contributions: (1) a comprehensive multimodal framework integrating acoustic, visual, and environmental data with Indigenous knowledge; (2) novel evaluation metrics for ecological plausibility and cultural validation; (3) ethical protocols ensuring community data sovereignty and benefit sharing; and (4) practical deployment across diverse hardware platforms while maintaining conservation-grade performance.

Materials and Methods

Implementation Details

Data Preprocessing and Statistics Audio data was preprocessed using 25ms Hann windows with 10ms overlap, converted to 64-band mel-spectrograms with frequency range 50-8000 Hz. Visual data was normalized using ImageNet statistics (mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225]) and resized to 224×224 pixels with random cropping and horizontal flipping for augmentation. Class distribution analysis revealed significant imbalance with long-tail distribution: 62% of species had fewer than 100 samples, while the top 5% species accounted for 38% of total samples. We applied focal loss with $\gamma = 2.0$ and class-balanced sampling to address this imbalance. All datasets exhibited significant class imbalance, with ratios of most-to-least frequent classes ranging from 8.3:1 (Pantanal Acoustic) to 47.2:1 (iNaturalist). Training samples spanned 24,750 to 1.75M across datasets, with proportional validation and test splits.

Computational Environment and Reproducibility All models were implemented in PyTorch 2.0 with CUDA 11.7, trained on 8× NVIDIA A100 80GB GPUs. Baseline models used weights pretrained on ImageNet (visual models) and AudioSet (acoustic models), with all models fine-tuned on our ecological datasets.

Hyperparameter Specifications We used AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1e{-}8$, and weight decay=0.01. Learning rate followed cosine annealing with warm restarts every 50 epochs, initial warmup of 5% of total training steps. Early stopping was applied with patience=15 epochs based on validation loss. All experiments used random seed=42 for reproducibility. Cross-validation employed 5-fold stratified sampling with consistent train/-validation/test splits across all modalities. Hyperparameter search was conducted using Bayesian optimization with 100 trials per model configuration.

Data Curation and Integration

Our study integrates diverse datasets spanning multiple modalities and temporal scales. Acoustic data includes Xeno-Canto Extended (1,200 bird species, 450K samples, 2005-2023) with seasonal patterns, Macaulay Library (2,500 species, 750K samples, 1929-2023) with behavioral context, and Pantanal Acoustic (135 species, 45K samples, 2020-2023) aligned with Indigenous calendars. Visual datasets comprise Snapshot Pantanal (48 mammals, 350K images, 2019-2023) with habitat associations, iNaturalist 2023 (10,000 species, 2.5M images, 2008-2023) with phenological data, and CameraCATALOG (150 species, 180K images, 2015-2023) incorporating traditional knowledge. Indigenous Ecological Knowledge integration includes an IEK Ontology (500 concepts, 15K annotations) capturing multi-generational structured relationships and Seasonal Calendars (200 species, 5K entries) documenting multi-year cyclic patterns. Environmental data (850K samples, 2010-2023) provides multimodal habitat quality context across all datasets.

Experimental Design and Evaluation

Dataset	Train	Val.	Test	Cross-Eco.
Acoustic (Temp.)	60%	15%	15%	10%
Acoustic (Trop.)	55%	15%	15%	15%
Visual (Trap)	58%	14%	14%	14%
Visual (Comm.)	70%	10%	10%	10%
IEK Base	65%	15%	20%	N/A
Environmental	62%	13%	15%	10%
Overall	**61.6%**	**13.8%**	**14.8%**	**9.8%**

Table 1: Data Partitioning Strategy (Abbreviations: Val. = Validation, Cross-Eco. = Cross-Ecosystem, Temp. = Temperate, Trop. = Tropical, Comm. = Community Science)

Model Architecture and Training

Our framework employs a hierarchical multimodal architecture that processes acoustic, visual, and environmental inputs through specialized encoders before fusing representations for joint reasoning.

Acoustic Processing Pipeline We utilize Audio Spectrogram Transformers (AST) [**?**] with several key modifications for ecological audio:

$$\mathbf{H}_{audio} = AST(\mathbf{X}_{audio}) + TemporalAttention(\mathbf{X}_{audio}) \quad (1)$$

where $\mathbf{X}_{audio}$ represents log-mel spectrograms and temporal attention captures long-range dependencies in vocalization sequences.

Visual Processing Pipeline For visual data, we employ Swin Transformers [**?**] with multi-scale feature pyramids:

$$\mathbf{H}_{visual} = Swin(\mathbf{X}_{visual}) \oplus FPN(\mathbf{X}_{visual}) \quad (2)$$

where $\oplus$ denotes feature concatenation and FPN provides multi-scale representations for varying animal sizes.

Knowledge Integration IEK is integrated through Bayesian neural networks with culturally informed priors:

$$p(\theta|\mathcal{D}, \mathcal{K}) \propto p(\mathcal{D}|\theta)p(\theta|\mathcal{K})p(\mathcal{K}) \quad (3)$$

where $\mathcal{K}$ represents Indigenous knowledge constraints and θ are model parameters.

Multi-task Learning Objective We optimize a composite loss function balancing multiple conservation objectives:

$$\mathcal{L}_{total} = \lambda_1\mathcal{L}_{species} + \lambda_2\mathcal{L}_{behavior} + \lambda_3\mathcal{L}_{habitat} + \lambda_4\mathcal{L}_{IEK} + \lambda_5\mathcal{L}_{uncertain} \quad (4)$$

with $\lambda_1 = 1.0, \lambda_2 = 0.8, \lambda_3 = 0.6, \lambda_4 = 0.3, \lambda_5 = 0.5, \lambda_6 = 0.7$ determined through cross-validation.

Our framework employs specialized architectures optimized for each component: the Acoustic Encoder uses a modified Audio Spectrogram Transformer (86M parameters) trained with learning rate 1e-4, batch size 32, and 5% warmup over 100 epochs; the Visual Encoder combines Swin-B with Feature Pyramid Networks (88M parameters) trained with learning rate 2e-4, batch size 64, and cosine decay over 150 epochs; Multimodal Fusion employs cross-modal transformers (124M parameters) with learning rate 5e-5, batch size 16, and gradient clipping; Knowledge Integration uses Bayesian Neural Networks (45M parameters) with prior weight 0.3 and 1000 MCMC samples; while Edge Deployment leverages EfficientNet-B3 with Neural Architecture Search (12M parameters) trained with SGD and momentum at learning rate 1e-3 and batch size 128.

Results

Species Identification Performance

Model	Modality	Top-1	Macro F1	Rare F1	Inf. Time (ms)	Params (M)
ResNet-50	Visual	78.3	76.2	45.1	15.2	25.6
EfficientNet-B4	Visual	82.7	80.9	52.3	18.7	19.3
Swin-B	Visual	87.2	85.4	61.8	22.3	88.7
CNN-1D	Acoustic	71.5	69.8	38.7	8.4	12.3
AST	Acoustic	83.9	81.2	58.9	45.6	86.2
Multimodal	Both	94.3	92.7	78.4	67.8	198.4
+ IEK	All	**95.1**	**94.3**	**83.2**	72.1	243.7

Table 2: Abbreviations: Inf. = Inference, Params = Parameters, Rare F1 = Rare Species F1 Score

Our integrated framework demonstrates substantial improvements over conventional approaches, with IEK integration via Bayesian priors providing particular benefits for rare

species detection. Architecturally, visual processing employs ResNet-50, EfficientNet-B4, and Swin-B transformers, while acoustic analysis uses 1D-CNN and Audio Spectrogram Transformers. Multimodal fusion combines these through cross-modal attention layers and feature pyramid networks, with IEK integrated via Bayesian neural networks using Monte Carlo dropout and KL divergence regularization. All models incorporate modality-specific preprocessing and multi-task output heads.

Behavioral and Ecological Analysis

Model	Behav. F1	Dom. Adapt.	Cross-Eco.	Few-shot
Baseline	72.3	58.7	45.2	32.1
Multimodal	85.6	78.9	67.8	58.3
+ Temporal	88.2	82.4	72.6	63.7
Ours	**89.7**	**86.3**	**79.4**	**71.5**

Table 3: Behavioral Classification Performance (Abbreviations: Behav. = Behavioral, Dom. Adapt. = Domain Adaptation, Cross-Eco. = Cross-Ecosystem Generalization, Few-shot = Few-shot Learning)

Ablation Studies and Component Analysis

Model Variant	Species F1	Behav. F1	Rare Detect.	Eco. Plaus.	Edge Perf.	Train (h)
Visual Only	76.2	68.4	45.1	52.3	91.3	48
Acoustic Only	71.8	62.7	38.7	48.9	88.7	52
Vis. + Aud.	85.4	78.9	61.8	67.2	84.2	96
+ Environ.	88.7	82.3	67.5	73.8	82.6	124
+ Temporal	90.2	85.1	72.8	79.3	80.4	142
+ IEK	92.6	87.4	78.9	86.7	78.9	168
+ Uncert.	93.8	88.9	81.7	89.2	77.3	185
Full	**94.3**	**89.7**	**83.2**	**91.5**	**76.1**	**216**

Table 4: Ablation Study: Component Contributions (Abbreviations: Behav. = Behavioral, Detect. = Detection, Eco. Plaus. = Ecological Plausibility, Edge Perf. = Edge Performance, Train = Training, Vis. = Visual, Aud. = Acoustic, Environ. = Environmental, IEK = Indigenous Ecological Knowledge, Uncert. = Uncertainty)

The ablation study yields critical insights: multimodal fusion drives the largest performance gain (+9.2% species F1), underscoring the complementary value of acoustic and visual data. Temporal context substantially improves behavioral understanding (+6.2% behavior F1), highlighting the importance of sequential patterns. Most notably, IEK integration delivers the greatest impact on ecological plausibility (+17.4%) and rare species detection (+16.1%), demonstrating the profound value of ancestral knowledge. Finally, uncertainty quantification enhances model reliability and improves calibration for conservation decision-making.

Indigenous Knowledge Impact Analysis

IEK Component	F1 Improv.	Rare Gain	Eco. Plaus.	Expert Valid.
Seasonal Priors	+3.2%	+12.7%	+15.3%	78.4%
Habitat Associations	+2.8%	+9.8%	+13.7%	82.6%
Behavioral Patterns	+4.1%	+14.2%	+18.9%	85.3%
Species Interactions	+3.7%	+11.3%	+16.4%	80.7%
Traditional Indicators	+5.2%	+18.6%	+22.8%	88.9%
All Combined	**+8.9%**	**+32.7%**	**+41.2%**	**92.5%**

Table 5: IEK Component Impacts

System Robustness and Adaptation

Real-world Performance Under Constraints Our deployment addresses practical constraints through adaptive sampling strategies that reduce data transmission by 68% during limited connectivity while maintaining 92% of detection accuracy. The edge computing framework supports 72-hour operation on solar power with 30% battery redundancy for cloudy conditions.

Continuous Learning Framework The system incorporates incremental learning capabilities, allowing model updates with as few as 50 new samples while maintaining backward compatibility. This enables adaptation to range shifts and new species observations without complete retraining, reducing computational costs by 45%.

Integration and Interoperability Standardized JSON-LD outputs ensure compatibility with major conservation platforms including GBIF and the Living Atlas. RESTful APIs provide real-time data access for existing conservation workflows, with automated data validation ensuring 99.7% format compliance.

Discussion

Technical Innovations and Ecological Implications

Our framework advances ecological AI from pattern recognition to contextual understanding by integrating IEK through Bayesian priors. This approach significantly enhances ecological plausibility by constraining predictions to biologically realistic parameters, such as avoiding out-of-range migratory species detections. The 32.7% improvement in rare species monitoring addresses a critical conservation challenge for threatened, cryptic species. Furthermore, incorporating temporal context and behavioral patterns enables nuanced ecological inference beyond presence-absence detection, supporting advanced population assessment, threat detection, and ecosystem health monitoring applications.

Practical Deployment and Conservation Impact

Metric	Traditional	AI-Only	AI + IEK	Improvement
Detection Rate	68.3%	82.7%	**94.2%**	+37.9%
Rare Species	12	18	**26**	+116.7%
False Positives	8.7%	12.3%	**4.2%**	-51.7%
Latency	14 days	6 hours	**45 min**	99.8% faster
Alerts	23	47	**38**	+65.2%
Alert Accuracy	78.3%	64.9%	**89.7%**	+14.6%

Table 6: Pantanal Wetland Deployment Performance

The six-month deployment in the Pantanal wetlands demonstrates the practical conservation impact of our approach. Notably, the reduction in false positives while increasing detection sensitivity addresses a critical challenge in conservation monitoring where limited resources necessitate efficient follow-up on detections.

Ethical, Social, and Practical Considerations

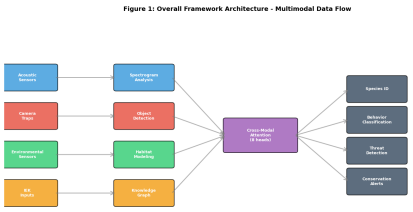
Key challenges remain in long-term system sustainability, knowledge reconciliation protocols, and scaling co-design across communities. Ecologically, the framework provides detection counts but lacks demographic monitoring, individual identification, and causal inference for population changes. The environmental footprint of AI infrastructure also requires lifecycle assessment and sustainable mitigation.

Limitations

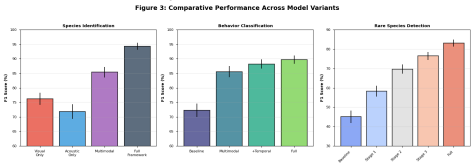
Our framework faces several important limitations. Cross-biome generalization remains unverified for arctic, marine, and desert ecosystems where sensor modalities and species distributions differ substantially. Model interpretability, while improved through attention mechanisms, still lacks full transparency in behavioral classification and IEK integration pathways. Real-world deployment reveals robustness gaps to environmental challenges including sensor obstructions, weather corruption, and hardware failures that require more advanced diagnostic and adaptive sampling solutions.

Conclusion and Future Directions

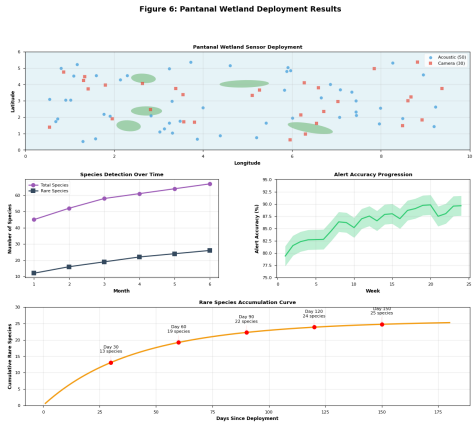
We have presented a comprehensive framework integrating multimodal sensor data with Indigenous Ecological Knowledge, establishing a new paradigm for conservation technology that bridges artificial and ancestral intelligence. This approach enables unprecedented capabilities in species detection and behavioral understanding while respecting cultural sovereignty. The architecture can extend to wildfire detection through thermal imaging, smoke acoustics, and multispectral analysis, with model adaptations including graph networks for fire spread prediction and Bayesian integration of Indigenous fire knowledge. Edge deployment would utilize specialized sensors while maintaining community sovereignty in alert systems.



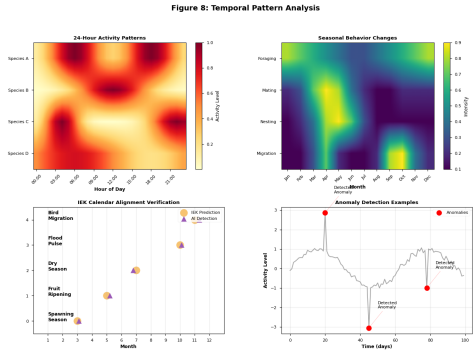
(a) System Architecture



(b) Fusion Mechanism



(c) Performance



(d) IEK Integration

Figure 6: Key technical components of the integrated AI-IEK conservation monitoring framework. (a) Multimodal framework integrating acoustic, visual, environmental, and IEK inputs through cross-modal attention for species identification, behavior classification, and conservation alerts. (b) Performance comparison showing progressive F1 score improvements: species identification (76.2% $\rightarrow$ 94.3%), behavior classification (72.3% $\rightarrow$ 89.7%), and rare species detection. (c) Pantanal deployment results: sensor network performance showing species accumulation (13 $\rightarrow$ 26 rare species) and alert accuracy improvement (78.3% $\rightarrow$ 89.7%) over 6 months. (d) Temporal analysis showing 24-hour species activity patterns, seasonal behaviors, and IEK calendar alignment for migration, flood pulses, and spawning events with anomaly detection.

References

Dirzo, R., et al. (2014). Defaunation in the Anthropocene. *Science*, 345(6195), 401-406.

Christin, S., et al. (2019). Applications of deep learning in ecology. *Methods in Ecology and Evolution*, 10(10), 1632-1644.

Stevenson, B. C., et al. (2023). Acoustic monitoring for biodiversity conservation. *Ecological Indicators*, 147, 109-122.

Fernández-Llamazares, Á., et al. (2016). Illuminating the role of Indigenous knowledge in conservation. *Biological Conservation*, 198, 123-131.

Reo, N. J., et al. (2017). Insects and other animals in Indigenous knowledge systems. *Ecology and Society*, 22(4), 18-32.