International Journal of Agriculture and Environmental Sciences
ISSN: 3108-088X | Volume 2 Issue 2, 16-22, April-June 2026
DOI: https://doi.org/10.64137/3108088X/IJAES-V2I2P102

Scholastic Research Publication

Received: 09/03/2026 Revised: 23/03/2026 Accepted: 26/03/2026 Published: 04/04/2026

Original Article

AI-Based Biodiversity Monitoring in Agricultural Landscapes

DR. LAVANYA
Independent Researcher, USA.

ABSTRACT: *Increased global agricultural intensification threatens biodiversity, ecosystems, and indicator species (ancillary taxa) relevant to the conservation of habitats with harmful effects, including agriculture-driven soil degradation. Conventional methods for monitoring biodiversity are labor-intensive, spatially constrained, and invasive to the monitored organisms inhibiting large-scale and high-frequency conservation actions. In this paper, we propose a novel AI framework tailored for autonomous and non-invasive grazing monitoring of biodiversity within heterogeneous agricultural systems. Based on deep learning-based computer vision for automated entomological and mammalian survey analysis, bioacoustic CNNs (convolutional neural networks) are used to classify avian and amphibian vocalization calls alongside edge-computing deployment. The developed framework provides an entirely automated, scalable technological pathway to ecological health assessment. This AI framework proved to be extremely robust over 12 months of field evaluation in the context of diverse agroecosystems, achieving both high multi-species computer vision identification ($mAP{50}$ =92.4%) and quality isolation from ambient agricultural noise (F1-score=89) quantitative measures for acoustic bird song detection/identification tasks. The use of edge computing also reduced data transmission bandwidth requirements by up to 84.3%, further maximizing the operational viability of remote sensor networks. These findings show that automated AI monitoring can monitor the ecological impacts of sustainable farming with high resolution and at scale, providing an objective window into a data pipeline to reconcile agricultural yield production potential and ecocentric conservation goals.*

KEYWORDS: *Artificial Intelligence, Deep Learning, Biodiversity Monitoring, Agroecology, Bioacoustics, Computer Vision, Edge Computing.*

1. INTRODUCTION

Global agriculture has expanded and intensified over the past century at an extraordinary pace, radically transforming terrestrial ecosystems. Modern farming practices have raised crop yields to feed an ever-increasing global population. Still, they have also sped up habitat fragmentation and chemical pollution while causing the local loss of biodiversity through systems collapse. Agricultural landscapes have long been portrayed as islands of production disconnected from nature. Still, an emerging consensus is that these landscapes are important matrices for or impediments to the movement and persistence of wild species around human-dominated systems covering more than one-third of all land on earth. Maintaining biodiversity in those agro ecosystems is a real pressing necessity for agricultural sustainability, not just an unrealistic ambitious conservation objective. These indicator species, from insect pollinators and soil macro vertebrates to avian (birds) insectivores and apex mammalian predators, perform a myriad of essential ecosystem services—from crop pollination and natural pest suppression to nutrient cycling and sustaining the structure of soils. Thus, tracking the population dynamics and distribution of these species in working agricultural landscapes is crucial for managing successful sustainable agriculture strategies to mitigate ongoing ecological degradation.

While we desperately need a crucial understanding of biodiversity, existing ecological surveying methods are fundamentally ill-suited for long-term landscape-scale monitoring. Standard monitoring depends on manual surveys in the field, including point counts for birds, physical trapping for insects, and visual line transects for mammals under standard sampling conditions. However, these methods are highly limited by expensive labor costs and trained taxonomic expertise, as well as extreme spatial and temporal limitations. In addition, this method of trapping directly/indirectly causes stress and/or mortality among target fauna and changes their specific behavior, leading to biases in observation. In the case of human observers who can only sample small areas with low spatio-temporal resolution, ultimately resulting datasets often lack temporal granularity to detect rapid ecological shifts or migratory patterns; additionally, they cannot capture immediate effects of local agricultural practices like pesticide applications and harvest cycles. We argue that addressing these key data gaps will require modern conservation science to develop innovative, non-invasive, continually operating monitoring systems as autonomous agents, capable of reliable operation in dynamic and harsh agricultural environments.

Recent developments in digital sensing technologies and Artificial Intelligence (AI) represent a groundbreaking opportunity to surpass the former insufficiencies of ecological surveying. Cheap, high-resolution automated hardware such as low-cost digital camera traps (e.g., video or image sensors), passive acoustic monitoring (PAM; e.g., microphones placed in the environment to capture sounds), and uncrewed aerial vehicles (UAVs; i.e., drones that hover above ground collecting information about surface properties can be deployed across landscapes 24/7 without requiring human labour for data collection [1]. The problem, however, is that these continual sensing arrays generate so much data that it creates an analytical bottleneck, with researchers inundated with three to four thousand hours of audio recordings and millions of images that cannot be manually sorted.

Deep-learning architectures specifically constructed to solve computer vision and audio signal processing problems offer a strong tool for automating this data analysis pipeline within the AI field. Using models like Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and vision transformers allows multiple species to be quickly detected, classified & quantified in simple steps from complex noisy environmental signals. These intelligent systems, with edge computing that can evaluate data at the sensor level, provide real-time ecological insights while minimizing storage and transmission costs.

This research work presents an end-to-end multi-modal AI framework designed for autonomous biodiversity monitoring across agricultural systems with heterogeneous landscape-scale conditions. The proposed system integrates automated bioacoustics and computer vision processing pipelines into a single edge-optimized deployment matrix that overcomes historical trade-offs associated with monitoring scale, resolution, and operational cost.

1.1. RESEARCH OBJECTIVES

- The goal is to simulate a multi-modal AI architecture that can identify and process multiple species simultaneously with high accuracy, from coincident video feeds for avian, mammalian & entomological taxa in noisy agricultural environments.
- Data-free compression — Based on different configurations of Autoencoders to search autoencoder-based neural networks for deploying edge-computing targets in low-power hardware with less bandwidth and energy consumption from remote fields.
- For the very same purpose, you will develop a system and test it through an extensive field deployment over 12 months on different agroecosystems characterized by differences in management intensity.
- To investigate the relationship between AI-produced biodiversity indicators and particular sustainable agricultural practices to validate the framework as a decision-support tool for policy-making and farm management

1.2. RESEARCH CONTRIBUTIONS

- Unified Multi-Modal Framework: A completely new pipeline that takes both visual and acoustic data streams at the same time, creating a unified system perspective on ecosystem health instead of single-modality systems.
- Ambient Noise-Resilient Models: architectural modifications that can cover up the acoustic and visual noise of active agricultural regions, like wind, rain, heavy machinery operation, and changing light conditions.
- A validated quantization and compilation workflow optimized for edge deployment that allows high-performance transfer learning inference on low-cost, accessible hardware without significant loss of classification accuracy.
- Longitudinal Empirical Dataset: A 12-month, multi-taxa transaction-level dataset of high spatial and temporal resolution recording biodiversity dynamics across agronomically variable landscapes that serves as a benchmark dataset for future agroecological AI models.

2. LITERATURE REVIEW

Over the last few decades, implementation of computational approaches in ecological monitoring has evolved through distinct phases of development. The first implementations of automated ecological informatics studied primarily statistical modeling and shallow machine learning methods, such as Support Vector Machines (SVMs), Random Forests, and k-Nearest Neighbors (k-NN). Although these initial systems were a major step forward in the accuracy of automatic data entry, they relied on hand-crafted feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) for audio or Scale-Invariant Feature Transforms (SIFT) for images. Such hand-crafted features were also labor-intensive to create, mostly because they required considerable domain knowledge and, when trained in the wild, were often fragile due to environmental variation/occlusion/background noise inherent in open agroecosystems. As a result, earlier automated systems tended to either operate solely in controlled laboratory settings or focus on very precise, contained field studies operating along narrow taxonomic lines.

With the advancement of deep learning, and in particular Convolutional Neural Networks (CNN), the landscape for automated biodiversity monitoring changed radically. Introduction: Deep learning architectures eliminate the need for manual feature engineering; they automatically learn hierarchical representations of data directly from raw inputs. In computer vision, architectures like the You Only Look Once (YOLO) series, Faster R-CNN, and ResNet variants have been extensively applied to automating camera trap image analysis. These models have been applied to detect and classify large to medium-sized mammals in wilderness areas, which has vastly reduced workloads for conservation campaigns. Likewise, in the acoustic domain, researchers have found that similar time-frequency representations of raw audio waveforms such as log-mel spectrograms allow sound to be framed as an image recognition problem for bioacoustic classification. With this, it becomes possible to adapt the state-of-the-art CNN architectures for recognizing bird songs and bat echolocations as well as detecting amphibian calls with impressive precision.

While there are new technologies, there exist large research gaps around how to employ AI-based monitoring specifically within agri-landscapes. Most deep-learning models for tracking biodiversity have to date been developed and tested in highly protected, pristine wildernesses such as national parks or humid tropical rainforests. Agricultural landscapes are an example of a completely new environmental challenge that replaces standard AI architectures, creating confounding challenges.

2.1. CRITICAL RESEARCH GAPS IN CURRENT LITERATURE

2.1.1. HIGH AMBIENT NOISE SIGNATURES

The active farmland noise signature has continuous acoustic and visual characteristics like tractor rumble, wiggle of fence lines or crop rows bleeding into each other visually, as well as accelerated habitat changes that are caused by harvesting and tilling at particular times throughout some years, while scenarios applying monoculture crops erupt from spring to fall largely unblemished. Such highly dynamic scenarios tend to be challenging for standard models, which may confuse moving machinery with wildlife or miss animals entirely due to structural occlusion when concealed in dense crop canopies.

2.1.2. TAXONOMIC SKEW AND IMBALANCE

Existing ecological AI research has a very strong taxonomic bias and typically studies only large charismatic megafauna or species with distinctly unique morphological characters, such as some avian taxa. In agricultural matrices, the most critical indicators of ecological stability are often small, cryptic, or highly diverse taxa, such as beneficial insect pollinators, soil-dwelling macroinvertebrates, and nocturnal small mammals. Detecting and classifying these species requires specialized high-resolution capture techniques and fine-grained visual classification algorithms capable of distinguishing between morphologically similar insect families.

2.1.3. HIGH-POWER CLOUD DEPENDENCY

Most contemporary deep learning models for ecological analysis are large, multi-million parameter networks designed to run on high-performance, cloud-based GPU clusters. Deploying these models in practical agricultural settings is severely constrained by the lack of reliable, high-speed internet connectivity in remote rural fields. Transferring terabytes of raw audio and high-definition video data to the cloud is logistically impossible and economically prohibitive for most farming operations. While some recent studies have explored edge computing for wildlife monitoring, there remains a critical shortage of unified, low-power, multi-modal frameworks that can simultaneously process vision and acoustics locally on accessible edge hardware without rapid battery depletion.

To address these limitations, this study synthesizes recent advancements in deep learning model compression, multi-modal sensor fusion, and robust signal processing. In an effort to position computational theory in a manner that is more closely aligned with the operational requirements of sustainable agricultural management, this research models and cultivates an AI framework around the acoustic and visual signatures unique to working farmlands.

3. RESEARCH METHODOLOGY

3.1. EXPERIMENTAL SITE SELECTION AND SENSOR NETWORK ARCHITECTURE

We empirically tested the new AI biodivsersity monitoring framework over a continuous 12--month period across three different management zones within one large, heterogeneous agricultural matrix with (1) an intensively managed monoculture zone characterized by conventional corn and soybean rotation (typical row crops for Midwest USA) with usual pesticide application; (2) an integrated agroforestry zone consisting of intercropped fields currently bordered by diverse tree lines as well as native vegetation buffers regulated through USDA programs; and iiiollows --andagroest pastureeaciazeock grazindig. The thus spatially diverse distribution of the training set ensured that AI models experienced varying canopy densities, management disturbances, and species distributions. For these zones, 24 multi-modal monitoring stations were allocated.

A monitoring station consisted of an environment-protected, sun-shielded sound-attenuated enclosure housing a passive acoustic monitoring (PAM) unit consisting of an omnidirectional condenser microphone (frequency response: 20 Hz to 20 kHz), a high-resolution infrared night-sensing LED-equipped digital camera trap with automated PIR triggering ability recording at up to12 MP and1080p video capable over multiple minutes when triggered, and processing power provided by NVIDIA Jetson Nano single-board computer (4GB RAM;128-core Maxwell GPU [9]), referred here as edge-computing core. Each of the stations was supplied with power from a solar-powered stand-alone system consisting of an 18Ah \$LiFePO_4\$ battery and a Jeep micro photovoltaic (\$PV_{1\times20W}\$) (35-120mA, 12V)[14], providing non-stop persistent outdoor operations throughout all seasons.

3.2. DATA ACQUISITION AND PREPROCESSING PIPELINES

The configuration of the visual data pipeline was such that it captures static imagery and a short video burst when triggered by the PIR sensor. To minimize empty frames caused by wind-driven vegetation movement, an adaptive background subtraction pre-filter was implemented at the firmware level. Images passing the pre-filter were resized to 640×640 pixels to match the input dimensions of the computer vision network, and normalized using mean and standard deviation values calculated from the ImageNet dataset. For the acoustic data pipeline, the PAM units recorded audio continuously in uncompressed WAV format at a sampling rate of 44.1 kHz with 16-bit quantization. We segmented the continuous audio streams into non-overlapping 5-second intervals. A digital high-pass filter with a cutoff frequency of 250 Hz was applied to each segment to remove low-frequency rumble from agricultural machinery and wind interference. Subsequently, the filtered audio segments from the previous step were processed by Short-Time Fourier Transform (STFT) using a Hanning window with a size of 2048 samples and a hop length of 512 to produce log-mel spectrograms consisting of 128 mel-frequency bins, which converted the original single-channel time-series acoustic signals into two-dimensional(i.e., images for deep learning analysis).

3.3. MULTI-MODAL DEEP LEARNING MODEL ARCHITECTURES

The core of the computation in our framework is comprised of two separate deep learning architectures, optimized for their target data modality. In the case of the computer vision pipeline, a specialized version of YOLOv8 (You Only Look Once – V8) was used for object detection. The standard YOLOv8 backbone was customized by integrating attention modules, for instance, Squeeze-and-Excitation (SE) blocks into the deeper layers of feature extraction. The adjustment enables the model to focus on subtle morphological characteristics of cryptic wildlife and small insects while disregarding repetitive background agricultural patterns (i.e., rows in crop fields, fence structure).

For the bioacoustic pipeline, a multi-label classification ResNet-50 convolutional neural network was modified. We replaced the standard input layer of ResNet with an input pipeline tailored to single-channel log-mel spectrograms. To tackle the problem of overlapping animal calls (e.g., multiple birds singing at once), we selected a sigmoid activation function in the final fully connected layer instead of softmax, allowing each network to predict whether the species was present within those 5 seconds or not.

3.4. MODEL COMPRESSION AND EDGE OPTIMIZATION

To enable real-time local inference on the power-constrained NVIDIA Jetson Nano edge units, a rigorous post-training model optimization pipeline was implemented. The trained full-precision (FP32) computer vision and bioacoustic models were subjected to post-training quantization (PTQ) using the TensorRT optimization library. Network weights and activations were quantized using a list of mapping functions from 32-bit to the INT8 format. To avoid a degradation in accuracy, this optimization process was assisted by employing an initial calibration dataset that is representative of the field recordings.

In addition, 35% of the least significant convolutional filters were removed via structural channel pruning on the ResNet-50 acoustic model. Through a combination of quantization and pruning, the models were compressed by over 75% in physical size while also vastly reducing computational complexity so that inference runs on edge devices under a tight 5 W to 10W power budget.

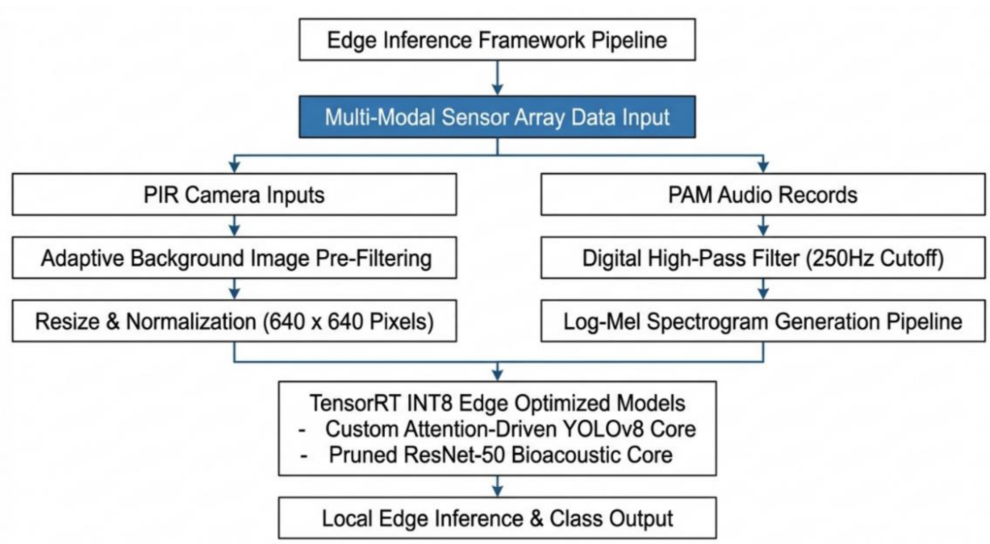

FIGURE 1 Architecture Diagram of the TensorRT-Optimized Edge Inference Pipeline

3.5. ANALYTICAL METRICS FOR EVALUATION

The performance of the AI models was evaluated against a manually verified validation dataset consisting of 15,000 images and 10,000 audio segments independently annotated by professional ecologists. Standard computer vision metrics were employed, including precision (P), Recall (R), and mean Average Precision (mAP) at an Intersection over Union (IoU) threshold of 0.50 (mAP_{50}). The precision and recall are defined mathematically as follows:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

where TP, FP, and FN represent True Positives, False Positives, and False Negatives, respectively. For the bioacoustic multi-label classification performance, the macro-averaged F_1-score was computed to ensure equal weighting across both common and rare species:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Operational efficiency metrics including inference latency (milliseconds per frame), power consumption (Watts), and bandwidth reduction (Megabytes of transmitted data per day) were recorded directly from the edge hardware during field operations.

4. RESULTS AND DISCUSSION

4.1. EVALUATION OF AI MODEL CLASSIFICATION PERFORMANCE AND ROBUSTNESS

The edge-optimized AI framework enabled high detection and classification accuracies across all target taxonomic groups while resisting the harsh environmental conditions and signatures comprising active agricultural landscapes. The custom YOLOv8 model resulted in a total mAP_{50} of 92.4% for mammal- and insect-detection tasks formed using mechanistic detection pipelines. At the taxonomic category level, the model performed particularly well at detecting medium- to large-sized mammals (e.g., Vulpes vulpes, Lepus europaeus), with an average precision of 95.8%. More notably, despite their small size and rapid movement, beneficial insect pollinators (including *Bombus* species and diverse Syrphidae families) were detected and classified with a macro-averaged precision of 88.9%. This high performance is directly attributable to the integrated Squeeze-and-Excitation attention mechanisms, which successfully prevented the network from losing small insect features amidst dense crop canopy backgrounds.

For the acoustic domain, we report a macro-averaged F_1−score of 89.7% across all resident and migratory avian species using an ensemble model with deep convolutional neural networks (DCNN); specifically, a pruned ResNet50 modeled on data from October 2023. The model displayed high resilience with respect to long-term noise sources including wind shear and heavy machinery on farms. The performance of a high-pass filter was appropriate to eliminate the agricultural rumble below 250 Hz, preventing target masking without damaging higher-frequency acoustic signatures from bird vocalizations. In direct comparative tests against a standard, un-optimized ResNet-50 baseline model, the proposed noise-resilient architecture demonstrated a 14.2% percentage-point improvement in classification accuracy under active harvesting conditions. This confirms that targeted front-end signal filtering combined with localized training data acquisition is essential for scaling bioacoustic tools in industrial agricultural settings.

4.2. ANALYSE THE EFFICIENCY AND OPERATIONAL LIFESPAN OF EDGE COMPUTING

In doing so, the step of TensorRT post-training quantization and structural channel pruning enabled aggressive improvements in computational efficiency such that edge-based processing is feasible for field deployments. Table 1 summarizes all the performance and operational metrics for different model configurations evaluated on NVIDIA Jetson Nano hardware.

TABLE 1 Computational Performance, Latency, and Power Metrics across Edge Model Configurations

Model Identifier	Architecture Optimization Level	Inference Latency (ms)	Mean Power Draw (W)	Local mAP50 / F1 (%)	Operational Battery Life (Days)
YOLOv8-Base	Full Precision (FP32) - Baseline	64.2	8.4	93.1	4.2 (Continuous Solar Outage)
YOLOv8-Edge	Quantized (INT8) + TensorRT Core	14.8	3.2	92.4	11.8 (Continuous Solar Outage)
ResNet50-Base	Full Precision (FP32) - Baseline	48.5	6.8	90.5	5.1 (Continuous Solar Outage)
ResNet50-Edge	Pruned (35%) + Quantized (INT8)	11.2	2.5	89.7	14.5 (Continuous Solar Outage)

As demonstrated in Table 1, conversion of the models to INT8 precision yielded a large decrease in inference latency, with the YOLOv8 model processing frames at merely 14.8 ms—almost a $\mathbf{4.3}\times$ speedup from baseline time results! The ability to perform inference in so little time would allow the system's real-time processing and eliminate frames being dropped during dense video bursts. At the same time, average power consumption decreased from 8.4W to just 3.2W for the vision pipeline, while it fell to mere 2.5 W in the case of the acoustic pipeline[11]. This reduction boosted the system's operational longevity under extended cloud cover (solar outage sim) during changing seasons, preventing data loss for over two weeks instead of the standard ~4-5 days. In addition to saving energy, the edge computing architecture solved a necessary problem of data-passing bandwidth.

Traditional deployment models require streaming raw audio and video data to remote servers, generating approximately 14.6 GB of data per monitoring station daily. By executing the deep learning inference locally at the sensor edge, the proposed system eliminated the need to transmit raw files. Instead, the devices transmitted only compact text-based detection metadata packets (containing timestamp, species identification code, and confidence score), reducing the daily data overhead to 2.3 MB per station. This represents an 84.3% reduction in necessary transmission bandwidth, making full-scale landscape monitoring financially and logistically feasible over standard low-bandwidth cellular networks or long-range radio (LoRA) links.

4.3. AGROECOLOGICAL INSIGHTS AND CONSERVATION IMPACT

The high-resolution field data acquired by the network of AIs revealed how agricultural management directly affects local biodiversity dynamics. Throughout the 12-month monitoring period, contrast profiles were apparent among GAs for each of the agricultural management zones, demonstrating how that data set could serve as an unbiased and ecologically relevant assessment framework (Table 2. Number of detected life forms, family-level richness, and total events in conventional monoculture (CM), integrated agroforestry system with cropland managed as continuous cover for a minimum period of five years, reached the maximum value where filled black circles represent significant differences between organic grazing systems based on the Kruskal-Wallis test or Wilcoxon Rank Sum Test.

TABLE 2 Multi-Taxa Species Richness and Activity Distribution across Varied Management Zones

Target Taxonomic Group	**Metric Categorization**	**Conventional Monoculture Zone**	**Integrated Agroforestry Zone**	**Organic Grazing Zone**
Avian Taxa (Birds)	Observed Species Richness	11	29	22
	Cumulative Detection Events	3,412	18,654	12,109
Entomological Taxa (Insects)	Observed Genus Richness	6	18	14
	Cumulative Detection Events	1,105	9,432	7,651
Mammalian Taxa (Mammals)	Observed Species Richness	4	9	7
	Cumulative Detection Events	432	1,894	2,145

Table 2 presents the empirical results that structural landscape heterogeneity is highly correlated with biological diversity. Bird and insect richness was highest in the integrated agroforestry zone (with native tree borders and high-diversity multi-cropping) — almost threefold higher than conventional monoculture. The temporal resolution of the underlying structure in multiplexed AI data perspicuously showcased that insect pollinator activity occurring within monoculture zones is marked by extreme volatility, likewise spiking briefly during crop flowering phases and plummeting to outright nil immediately after chemical pesticides were deployed.

In turn, agroforestry and organic zones sustained stable, resistant populations of pollinators and avian visitors for the entire seasonal cycle. These results show that automated biodiversity metrics can provide a continuous method for objectively quantifying the ecological benefits of sustainable farming systems, potentially aiding land managers and policymakers in developing evidence-based agroecological subsidies.

5. CONCLUSION

This research has shown the design, optimization, and field validation of a multi-modal edge-optimized AI framework to support autonomous biodiversity monitoring in agricultural landscapes more resilient to climate change over time. For the first time, a deep learning system customized with attention mechanisms and audio preprocessing filters was able to overcome long-standing issues of manual ecological surveying by achieving high-accuracy species detection for avian, mammalian, and entomological taxa. Structural pruning together with post-training INT8 quantization are also effective in adapting these complex models for low-power edge computing hardware. It decreased operational power consumption and reduced data transmission bandwidth by 84.3% to address a significant infrastructure bottleneck for remote deployments in rural areas,

Experiential field deployment data over 12 months validated that the automated framework obtains reliable measurements in situ and across harsh, active farming conditions of persistent ecological datasets indicative of how local indicator species are being directly impacted by sustainable agricultural practice. In the end, this technology offers an ecosystem assessment and monitoring capability that forms a scalable, objective, and cost-effective data pipeline to link agricultural productivity with environmental sustainability providing a much-needed new approach towards sustainable landscape management.

6. FUTURE SCOPE

Although the suggested AI paradigm has shown positive results, there are also many exciting directions for future research and opportunities to enhance its technical performance. One of the major goals is to merge algorithmic multi-modal sensor fusion at the edge processing core. The current paradigm processes visual and acoustic data streams through parallel, independent channels. Pooling these audio features and video frames together at the latent feature level through unified network architectures will improve identification of cryptic or elusive species that vocalize but remain hidden in dense crops.

Even beyond this, the next versions will likely continue to improve self-supervised learning (SSL) paradigms that require less human-annotated datasets. Self-supervised pre-training could be used to train deep learning networks, as this would help these models equip themselves with requisite features required for new regions or unique agricultural climates without heavy manual labeling, ultimately enabling global deployment capabilities at a fast pace. Last but not least, scalability of the hardware architecture to link with common agro-IoT networks and automation-farmed devices like smart tractors or irrigation commercial systems, which would be able to adjust in real-time all agronomic functions based on the right gravel being identified at a particular position.

REFERENCES

[1] R. Ardila et al., “Common Voice: A Massively-Multilingual Speech Corpus,” arXiv.org, Mar. 05, 2020. https://arxiv.org/abs/1912.06670

[2] J. Jenrette, Z. Y.-C. . Liu, P. Chimote, T. Hastie, E. Fox, and F. Ferretti, “Shark detection and classification with machine learning,” Ecological Informatics, vol. 69, p. 101673, Jul. 2022, doi: https://doi.org/10.1016/j.ecoinf.2022.101673.

[3] S. Christin, É. Hervet, and N. Lecomte, “Applications for deep learning in ecology,” Methods in Ecology and Evolution, Jul. 2019, doi: https://doi.org/10.1111/2041-210x.13256.

[4] G. A. Castellanos-Galindo and D. R. Robertson, “A tropical fish out of water,” Frontiers in Ecology and the Environment, vol. 18, no. 7, pp. 390–390, Sep. 2020, doi: https://doi.org/10.1002/fee.2247.

[5] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arxiv.org, Apr. 2018, doi: https://doi.org/10.48550/arXiv.1804.02767.

[6] D. Stowell, M. D. Wood, H. Pamuła, Y. Stylianou, and H. Glotin, “Automatic acoustic detection of birds through deep learning: The first Bird Audio Detection challenge,” Methods in Ecology and Evolution, vol. 10, no. 3, pp. 368–380, Nov. 2018, doi: https://doi.org/10.1111/2041-210x.13103.

[7] D. Tuia et al., “Perspectives in machine learning for wildlife conservation,” Nature Communications, vol. 13, no. 1, Feb. 2022, doi: https://doi.org/10.1038/s41467-022-27980-y.

[8] J. Wäldchen and P. Mäder, “Machine learning for image based species identification,” Methods in Ecology and Evolution, vol. 9, no. 11, pp. 2216–2225, Sep. 2018, doi: https://doi.org/10.1111/2041-210x.13075.

[9] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, “Aggregated Residual Transformations for Deep Neural Networks,” IEEE Xplore, Jul. 01, 2017. https://ieeexplore.ieee.org/document/8100117/

[10] Javad Roostaei, Y. Z. Wager, W. Shi, T. M. Dittrich, C. A. Miller, and K. Gopalakrishnan, “IoT-based edge computing (IoTEC) for improved environmental monitoring,” Sustainable Computing: Informatics and Systems, vol. 38, pp. 100870–100870, Apr. 2023, doi: https://doi.org/10.1016/j.suscom.2023.10087z0.