

Scaling up Multi-Sensor Monitoring of Forest Restoration with Computer Vision and Deep Learning: A Case Study from Hawai'i

Yuyan Chen*
McGill University
yuyan.chen2
@mail.mcgill.ca

Valentin Gabeff*
EPFL
valentin.gabeff
@epfl.ch

Maximiliane Jousse*
McGill University
maximiliane.jousse
@mail.mcgill.ca

Fedor Zolotarev*
University of Helsinki
fedor.zolotarev
@helsinki.fi

Ekaterina Nepovninnykh
LUT University
ekaterina.nepovninnykh
@lut.fi

Mike Long
National Ecological Observatory Network,
Battelle
mlong@battelleecology.org

Daniel Rubenstein
Princeton University
dir@princeton.edu

Abstract

Biodiversity monitoring is essential to evaluate whether restoration and conservation initiatives effectively mitigate biodiversity loss driven by climate change and habitat degradation. Given the complexity of ecological dynamics at play, monitoring the different variables of interest over time and at different spatial scales is challenging. In this study, we present a cost-effective and scalable framework to use multiple sensors for data collection and automate data processing with various computer vision and deep learning methods to facilitate long-term, large-scale monitoring of biodiversity. We illustrate the potential of this approach through a case study, in which we monitor a koa restoration project in the Pu'u Maka'ala Natural Area Reserve, Hawai'i. We study the relationship between maturity of restored patches and biodiversity variables by integrating information from camera traps, passive acoustic microphones, LiDAR and drone flights. While we process camera-trap and acoustic data semi-automatically with existing deep learning models, we also adapt active contour methods to measure koa forest growth from LiDAR and drone data. Specifically, we segment individual trees by first estimating ground and top surfaces from the LiDAR point clouds, and then by clustering elements from the top surface based on height density using watershed. We find an evidence of association between biodiversity from camera traps and microphones and koa tree maturity and demonstrate the effectiveness of using multi-sensor data and computer vision for biodiversity monitoring.

1. Introduction

Given the widespread need for conservation action worldwide, it is essential to monitor both past and ongoing conservation and restoration efforts to evaluate their effectiveness [23, 28]. Insights from such monitoring can inform future policies and strategies across various ecosystems, spatial and temporal scales, and decision-making levels. Autonomous data collection systems, including camera traps, passive acoustic microphones (PAM), drones, and LiDAR, enable large-scale, long-term, and multimodal (encompassing multiple modalities, i.e. different data types) ecosystem monitoring [2]. To leverage the large amount of data collected by such systems, computational methods are increasingly used to automate data processing [2, 22]. These systems significantly reduce the time and resources required to collect and process large amounts of data compared to traditional methods, while allowing for different dimensions of ecosystems to be monitored and the collection of new sources of data [22]. These data can therefore be leveraged to offer novel insights and increased understanding of ecosystems and their dynamics before, during, and after conservation efforts.

In this study, we explore the potential of autonomous and multi-sensor monitoring to evaluate conservation efforts of koa restoration sites in Pu'u Maka'ala Natural Area Reserve (PUUM). Located at an elevation of 1,600 meters, PUUM serves as a vital refuge for native bird species whose populations are undergoing severe declines due to habitat loss, climate change, introduced predators, and diseases, especially avian malaria spread by invasive mosquitoes [4, 7, 25, 32]. However, much of PUUM's area was previously converted to pasture. Introduced grasses dominate these areas and will continue to do so without human intervention [26]. Recent

*Equal contribution.

restoration efforts have focused on planting koa (*Acacia koa*), an endemic keystone species, to restore the ecosystem and provide habitat for endangered bird species [30]. Given the large area of the restored plot, we develop automated monitoring systems to record and process images and sounds of animals across plots of varying developmental stages to scale up and facilitate the analysis of the complex biodiversity dynamics associated with koa restoration over time.

We use foundation models to obtain species occurrence from camera trap and acoustics data. We also propose a novel method to measure tree growth using LiDAR and drone imagery. We then evaluate restoration success by linking extracted biodiversity variables to sites at different restoration stages and conclude that bird biodiversity increases as koa tree restoration plots mature in Pu'u Maka'ala. Our findings demonstrate the potential of using multimodal data along with automatic processing pipelines to assess and inform conservation efforts.

2. Data collection

We selected three sites within PUUM where koa have been replanted, named Open Grassland (OG), Park Land (PL) and Closed Canopy (CC), respectively (see Fig. 1). Although the exact planting dates of the restored koa stands are unavailable, preliminary observations based on koa tree size and grassland cover suggest that OG represents the most recently established site, CC the oldest, and PL an intermediate stage.

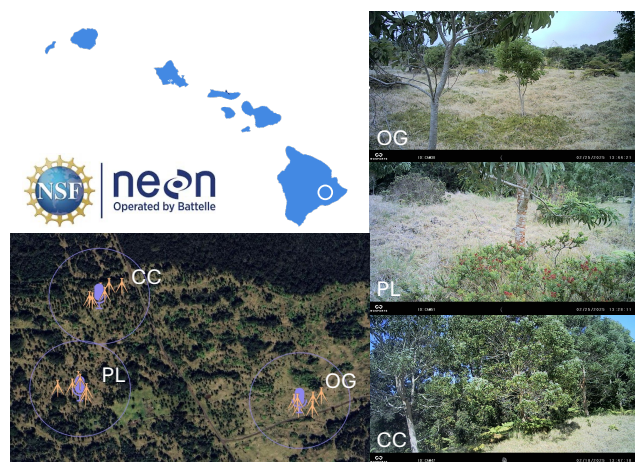


Figure 1. **Overview of the study sites and sensor placement.** Top left: map of Hawai'i archipelago with location of PUUM on Big Island Hawai'i (the white circle). Bottom left: visualization of the camera trap (orange) and microphone placement (purple). Right handed images obtained from one of the six camera traps placed at each site. From top to bottom: Open Grassland (OG), Park Land (PL), and Closed Canopy (CC).

We chose 6 focal koa trees with understory at each site. One camera trap was set up to point towards each focal tree with the intent to capture bird occurrences. All camera traps are motion-triggered, and were set to capture images with a burst of 3 images with a cool-down period of 5 seconds. We also implemented PAM by deploying one microphone at each site (CC, OG, PL, see figure 1). A tree close to the camera trap deployments (but not with a camera trap pointed at it) was selected and a microphone was fixed to a branch. Microphones were set to record for 30 minutes from 6:00 to 7:00 and 15 minutes every hour from 7:00 to 19:00. Finally, we flew a DJI Majic Mini drone over each site to measure variables associated with the structure of the koa stand. The drone was manually maneuvered, flying at an elevation of 20m, a camera angle of 60 degrees, and recording the scene at 30 frames per second.

3. Methods

3.1. Camera traps

We developed an automated pipeline to extract animal occurrence data using MegaDetector [11] and BioCLIP [29]. MegaDetector is a generalizable animal detector that can be used to filter out empty camera trap images due to false triggers. It has been widely applied to camera traps deployed in various ecosystems with varying degrees of success (e.g., [8, 16, 20]). BioCLIP [29], a recent foundation model for species recognition, leverages the potential of large vision-language models to achieve zero-shot species recognition in different geographic regions.

We first applied MegaDetector [1, 11] to all camera trap images to detect animal occurrences and filter out empty frames. MegaDetector contains three categories: animal, person, and vehicle. All detections labeled as "animal" were subsequently passed to BioCLIP [29] for species recognition. We compared the off-the-shelf performance of BioCLIP with the performance of classifiers trained with features extracted by BioCLIP.

We also developed a filtering algorithm to eliminate redundant detections caused by: 1) the camera trap setting (three images captured per trigger), 2) animals staying in the same location for extended periods, and (3) leaves consistently misclassified as birds over time by the species classifier. The algorithm groups detections classified as bird using their timestamps and bounding box coordinates. Specifically, we compute the Intersection over Union (IoU) between bounding boxes, grouping those with an IoU greater than 0.9. Within each group, we calculate the detection duration based on timestamps. If the duration is longer than 10 minutes, we consider the detections as non-bird, as such prolonged detections are likely to be misclassified leaves instead of a bird.

3.2. Bioacoustics

Raw audio recordings were split into 5-second clips to standardize analysis. Each clip was processed using an unsupervised sound source separation model based on Mixture Invariant Training (MixIT) ([5, 33]). Each extracted source was passed through a detector and sources classified as pure noise were discarded. The non-noise sources were passed to Perch, a bird species classifier that assigns probabilities to different bird species [9]. Along with classification labels and probabilities, Perch also produces a vector representation, or embedding, of the audio input.

Total species richness from PAM was calculated by counting the number of species detected at each site over the whole time period for samples that had a prediction confidence threshold above 0.95 and that occurred at least 50 times. We also computed a measure of diversity in the embedding space by running a principal component analysis (PCA) and the sourced embeddings for samples that had a confidence prediction threshold above 0.01, and reporting the number of principal components required for certain percentage of explained variance. The advantage of this method is that it is a measure of statistical diversity that is not biased by the accuracy of the bird species classification model while remaining related to a measure of biodiversity.

3.3. Analysis of biodiversity data

Results from camera traps and PAM were analyzed separately. We ran a Chi-squared test on the occurrence counts of species grouped per site to investigate whether a relationship was present between sites and occurrences. We then calculated the Shannon-Wiener Diversity Index (H), a metric of species diversity that takes into account both species richness and evenness [27]. From this metric, we calculated evenness (eH) separately. Finally, we compared among-site diversity using the Jaccard [13] and Bray-Curtis [3] dissimilarity metrics, the first on presence-absence and the second on occurrence data. All metrics were calculated in R using the package *vegan* [18]. Evenness is defined as $eH = \frac{H}{\ln(S)}$, where H is the Shannon-Wiener index at a site and S is the total number of species at that site [21].

3.4. LiDAR

In order to process LiDAR data of the selected plots of land, the point cloud was first rotated such that x , y and z axes are aligned with the first 3 PCA directions, eliminating natural trends due to elevation and other possible factors. Then, a modification of the classical active contour [15] algorithm was used to estimate the surface of the ground and the top surface. Active contours have been previously applied to LiDAR data for surface reconstruction [31] and forest terrain modeling [24], typically using energy functions that balance surface smoothness with point cloud fidelity. Our approach adapts this framework by incorporating a specific RBF ker-

nel energy formulation that uses distance-weighted point attraction for both ground and canopy surface estimation. The surface is represented as a grid of points with fixed x and y coordinates, and the goal is to optimize z coordinates by using energy based on distance to the lidar points. The point energy is defined as an RBF kernel function

$$E_{\text{point}}(s) = \sum_{p \in \text{KNN}(s)} \frac{1}{1 + \epsilon \|s - p\|^2}, \quad (1)$$

where s is the surface point, p is a LiDAR point from the set $\text{KNN}(s)$ of closest k points to the s , and ϵ is the shape parameter controlling the radius of attraction. This formulation differs from traditional active contour approaches that typically use Gaussian or polynomial energy functions. To increase interpretability of the parameters, the range parameter ϵ is defined as:

$$\epsilon = \frac{1 - t}{tr^2}, \quad (2)$$

where t is the threshold and r is the radius. With this ϵ , all points closer than r will have energy higher than t . The optimization was done as usual by moving z coordinates of surface points based on the gradient of the point energy. Ground surface was initialized with the minimum z coordinate from all points. Top surface was initialized with the maximum z coordinate. Intuitively, the top surface can be compared to dropping a blanket on top of the scene, where its stretchability and stiffness are controlled by active surface parameters α and β respectively.

With ground and top surfaces estimated, it is possible to get normalized information about tree heights by subtracting ground z from all points and top surface. By treating top surface as an image with intensities corresponding to heights, it is possible to apply watershed segmentation to separate objects of different sizes. Watershed segmentation is widely used for individual tree detection from LiDAR-derived height models [34, 35], and we apply this standard approach to our refined surface estimates. Since canopies tend to be somewhat spherical, the valley between overlapping canopies will get separated by the watershed segmentation, resulting in a more granular result. To get final trees, LiDAR points were thresholded based on the distance from the ground and assigned an ID based on the segmentation result.

4. Results

4.1. Camera traps

We first evaluated the off-the-shelf performance of BioCLIP on the camera trap data. We manually annotated 3,358 image crops that were identified as "bird" by BioCLIP. The resulting dataset includes seven bird species and a "non-bird" category (see Appendix Table 9). Of the annotated crops, 74% are non-bird. For the remaining 875 bird-labeled crops, BioCLIP achieved a species-level accuracy

of 8.65%. Therefore, instead of directly using BioCLIP’s predictions, we trained a classifier with features extracted by the model.

Method	Training set	Accuracy	Macro F1
KNN	Crop	98.11	41.35
	Crop + iNaturalist	97.91	49.14
Ridge	Crop	96.52	26.84
	Crop + iNaturalist	96.42	60.21

Table 1. **Ablation study on training data and methods.** The highest accuracy and Macro F1 scores are **bold**.

To assess the accuracy of this approach, we split the annotated dataset into training and testing, and further augmented the training set with images downloaded from iNaturalist¹ (details in the Appendix B). We did an ablation study to see whether using iNat images improve the performance and whether K-Nearest Neighbors (KNN) [10, 19] or Ridge Classifier [12, 19] perform better on this task. Since our dataset is highly imbalanced, we used both accuracy and Macro F1 score as evaluation metrics. As shown in Table 1, all methods achieve high accuracy and the Ridge classifier trained with both iNat and camera trap images achieve the highest macro F1 score. Hence, we used the Ridge Classifier trained with both camera trap and iNaturalist data to process camera trap data.

We applied the processing pipeline to data collected by 18 camera traps from January 23rd to May 28th, 2025 (125 days) to obtain species occurrence in the three sites. We summarize the results in Table 2. Scientific names of the species can be found in the Appendix (Table 8).

We performed Pearson’s Chi-squared test on the results presented in Table 2 to test whether species occurrences are independent of site. The p-value is $\ll 0.01$, suggesting a relationship between species occurrences and sites. More introduced species are observed in OG compared to PL or CC. In contrast, more native birds are detected in PL and CC, where they are seen perching or feeding on native Hawaiian plants such as koa, pilo, and 'ōhelo 'ai (see Fig. 2). Number of times recorded for each species also varies widely between species and sites.

Further analysis using the Shannon Diversity metric shows that PL has a higher diversity, followed by CC (see Table 2.). OG has a much lower biodiversity, in part explained by its very low evenness, driven by the nēnē, the most frequently detected species overall and in OG. Comparing sites to each other using dissimilarity metrics (see Table 3) shows that CC-PL are more similar to each other than CC-OG and PL-OG. Bray-Curtis dissimilarity further shows a higher dissimilarity between OG-PL than OG-CC.

¹<https://www.inaturalist.org/>

Common Name	OG	PL	CC
Nēnē	566	6	19
Hawai'i 'amakihi	0	0	2
Common waxbill*	1	0	0
'Apapane	0	3	6
Kalij pheasant*	4	17	5
'Ōma'o	0	6	0
American barn owl*	2	0	0
Species Richness	4	4	4
Shannon Diversity	0.08	1.19	1.09
Evenness	0.06	0.86	0.78

Table 2. **Predicted species observed by camera traps in three restoration sites (Open Grassland (OG), Parkland (PL), and Closed Canopy (CC)).** Values indicate the number of detections per species at each site from Jan. 23rd to May 28th, 2025 (125 days). Asterisks (*) denote non-native species. Species richness, Shannon-Weiner index and Shannon Evenness metrics for camera trap data at each site.

Jaccard	OG	PL	CC
OG	-	-	-
PL	0.67	-	-
CC	0.67	0.40	-
Bray-Curtis	OG	PL	CC
OG	-	-	-
PL	0.97	-	-
CC	0.92	0.56	-

Table 3. **Dissimilarity metrics (Jaccard and Bray-Curtis) between the three sites for the camera trap data.** Jaccard dissimilarity is based on presence-absence data, while Bray-Curtis dissimilarity is based on occurrence data.

4.2. Bioacoustics

We first filtered detections by prediction confidence, including only those with confidence scores greater than or equal to 0.95. For each site, we then summed the detections per species. Species with at least 10 detections at any site are reported in Table 4 (see Table 8 in the Appendix for scientific names). We performed a similar Chi-squared test as for the camera traps and obtained a Chi-squared value of 8567 ($p \ll 0.01$), also indicating that there is a relationship between sites and species occurrences. The Shannon Diversity Index further shows that higher biodiversity was observed at CC, closely followed by PL. Evenness reflects the same trend, with OG being the most uneven site, followed by PL and CC.

Both dissimilarity metrics show that CC-OG are the most dissimilar sites (see Table 5). However, the choice of using presence-absence or occurrence data to compute dissimilarity affects how dissimilar OG-PL are. OG-PL appear the most similar under Bray-Curtis dissimilarity, whereas

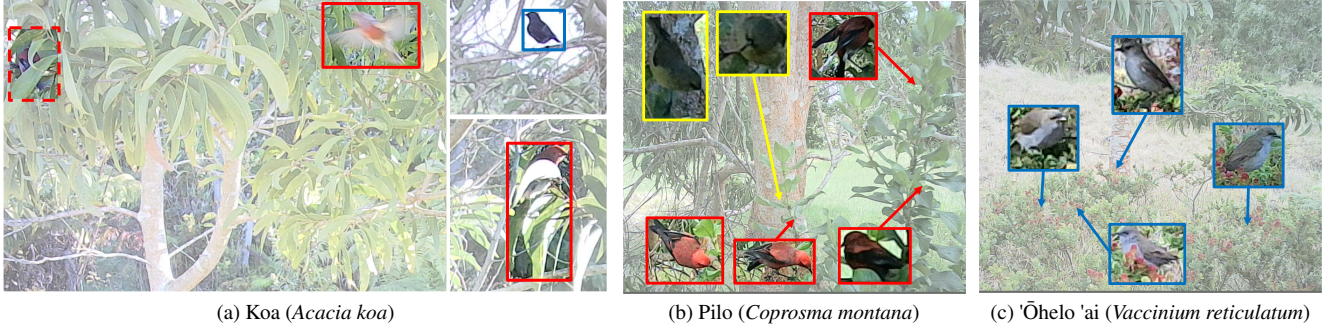


Figure 2. **Examples of native forest birds interacting with koa trees or plants (e.g. Pilo and 'Ōhelo 'ai) in their understory.** Solid bounding boxes represent detections by MegaDetector: red for 'Apapane, yellow for 'Amakihi, and blue for 'Ōma'o. The dashed red box shows an 'Apapane missed by MegaDetector due to occlusion.

Species	OG	PL	CC
'Akiapōlā'au	0	21	233
'Apapane	280	1964	4748
Cackling Goose*	105	10	4
Common Waxbill*	55	60	35
Hawai'i 'amakihi	2	24	44
Hawai'i Creeper	0	10	6
Nēnē	19	1	1
Japanese Bush Warbler*	499	868	3125
Northern Cardinal*	40	91	15
'Ōma'o	6142	5104	3183
Short-eared Owl	152	0	2
Spotted Dove*	67	2	0
Warbling White-eye*	424	1036	749
Wild Turkey*	2	78	43
Yellow-fronted Canary*	24	155	86
Species Richness	13	14	14
Shannon Diversity	0.92	1.35	1.42
Evenness	0.36	0.51	0.54

Table 4. **Predicted species observed across three koa sites (Open Grassland (OG), Parkland (PL), and Closed Canopy (CC)) derived from passive acoustic monitoring.** Values indicate the number of detections per species at each site from Jan. 23rd to July 9th, 2025. Asterisks (*) denote introduced species. Species richness, Shannon-Weiner index and Shannon Evenness metrics for Acoustic data (threshold 0.95) at each site.

CC-PL are the most similar according to the Jaccard index.

Embedding diversity from PCA measures are visualized in Figure 3 and described in Table 6. We observe that progressively more principal components are required to explain the same level of bird vocalization embeddings variance as the koa plots mature. Alternatively, a fixed number of principal components explains less vocalization embeddings variance in more mature sites. For example, principal components 1 to 5 explain 32.6 %, 37.6 % and 40.0 % of the variance for CC, PL and OG, respectively. This result

Jaccard	OG	PL	CC
OG	-	-	-
PL	0.20	-	-
CC	0.20	0.13	-
Bray-Curtis	OG	PL	CC
OG	-	-	-
PL	0.25	-	-
CC	0.56	0.36	-

Table 5. **Dissimilarity metrics (Jaccard and Bray-Curtis) between the three sites for the Passive Acoustic Monitoring data (threshold 0.95).** Jaccard dissimilarity is based on presence-absence data, while Bray-Curtis dissimilarity is based on occurrence data.

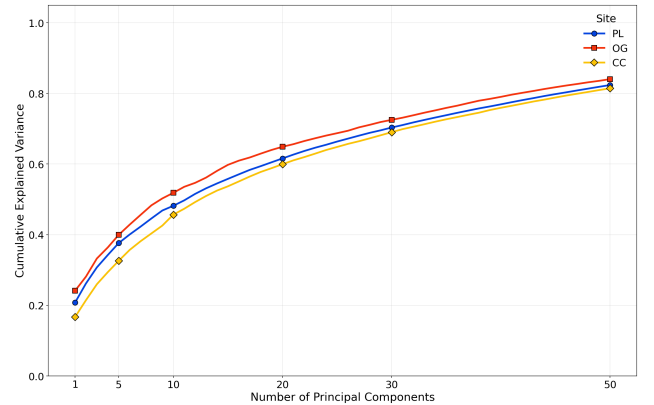


Figure 3. **Explained source separated vocalization embeddings variance per principal component for each site.** A site exhibiting a higher level of explained variance for a given number of principal components suggests that the vocalization variations at this site are less diverse than for other sites (Best viewed in colors).

suggests greater vocalization diversity in more mature sites, and overall greater total variance in more mature sites.

Site	N samples	Total variance	PCs _{80%}	PCs _{90%}	PCs _{95%}	EV PC1	EV PC1-5	EV PC1-10	EV PC1-20
OG	256470	5.287	42	66	81	24.11%	39.95%	51.85%	64.89%
PL	510747	5.330	46	68	83	20.76%	37.62%	48.17%	61.59%
CC	516497	5.688	48	70	84	16.68%	32.60%	45.67%	59.95%

Table 6. **Description of explained variance from PCA for the source separated sounds of each site.** EV: Explained Variance; PCs: Number of Principal Components. PCs_{n%} means the number of principal components to explain *n%* of variance. EV PC1-*k* means the percentage of variance that can be explained by the first *k* principal components.

4.3. LiDAR

The output of the koa trees segmentation algorithm from the LiDAR data are shown in Figure 4. We visually observe an increase in the number of individual koa trees at each site between 2019 and 2025 as well as an overall expansion of the canopy. These results are supported by the quantitative analysis of the segmentation output in Table 7. For all sites, we observe an increase in median koa tree height and number of individual trees over the 6 years period, representative of the restoration process.

While we also observe an increase in crown area and volume in PL and CC as expected for a growing forest, this is not observed for OG. Since the metrics are aggregated over the number of trees and many small koa trees were planted in OG between 2019 and 2025, the median crown diameter and volume is likely to be dominated by the dimensions of the new planted trees. This intuition is supported by the visualization of the segmentation results where the initial segmented elements (potentially not koa trees) were characterized by a relatively large area in comparison to the other elements detected in 2025.

When comparing sites among themselves, we observe that OG has the lowest statistics overall, followed by PL while CC exhibits the largest median height and crown volume. The quantitative metrics derived from the segmentation process are as expected and in line with the visual examination of each site (Figure 4) and field observations.

5. Discussion

5.1. Integrating multi-sensory information

The measures of plot maturity from our LiDAR processing approach confirm our intuitions about the maturity of each stand. CC and PL have a higher tree crown volume and height in comparison to OG, indicating more mature trees. Visual examination of the segmentation results from the 2019 and 2025 LiDAR surveys also confirms this result and aligns with our field observations.

The derived biodiversity measures from camera traps and passive acoustic microphones are consistent with each other. While species richness measured from camera traps are equal between sites (all sites recorded 4 species), we ob-

serve more evenness and diversity in more mature sites (CC and PL) than in presumably younger plots (OG). Similarly, the Jaccard index shows the strongest measure of dissimilarity between OG and the two other sites for both camera trap and acoustic data. While this result also holds for the Bray-Curtis index in camera traps, OG and PL show a strong similarity in biodiversity when compared with the Bray-Curtis index derived from acoustic data, likely driven by the large occurrences and relatively small differences between sites for those (e.g., 'Ōma'o). Finally, the statistical measure of diversity from the PCA analysis continue to support our hypothesis, where CC has the highest total variance of vocalization embeddings from Perch, while OG has the lowest. Moreover, more principal components are required to explain a fixed level of variance in more mature sites.

Taken together, the results derived by integrating information from all sensors (LiDAR, camera traps, PAM) indicate that we observe an increase in the associated bird biodiversity as koa tree restoration plots mature in Pu'u Maka'ala. As shown in this study, combining multiple sensors has the potential to derive novel site-level information at relatively little cost. With such sensors, data here-to-for widely inaccessible, such as acoustic or visual, can be integrated with other data to provide a picture of a landscape at a larger spatial and temporal scale. Furthermore these sensors are promising for monitoring in remote and vulnerable areas, as they are less disruptive than other monitoring methods.

5.2. Limitations

Despite the fact that our observations align with our hypothesis, however, our conclusions must be taken with care. Further ecological inference is not possible as our data collection protocol and processing pipelines contain a number of limitations and biases, outlined below, which prevent ecological conclusions from being drawn from the data.

Sensor placement. First, our data would be less biased given more sensors, but limited availability led to six cameras being placed in three different sites. While cameras at all sites were pointed at a "focal tree", bird species detected and the respective number of detections can still be affected by the choice of focal tree, the associated understory, the camera-view angle, and several other variables affecting detection probability. Furthermore, equipment limitations

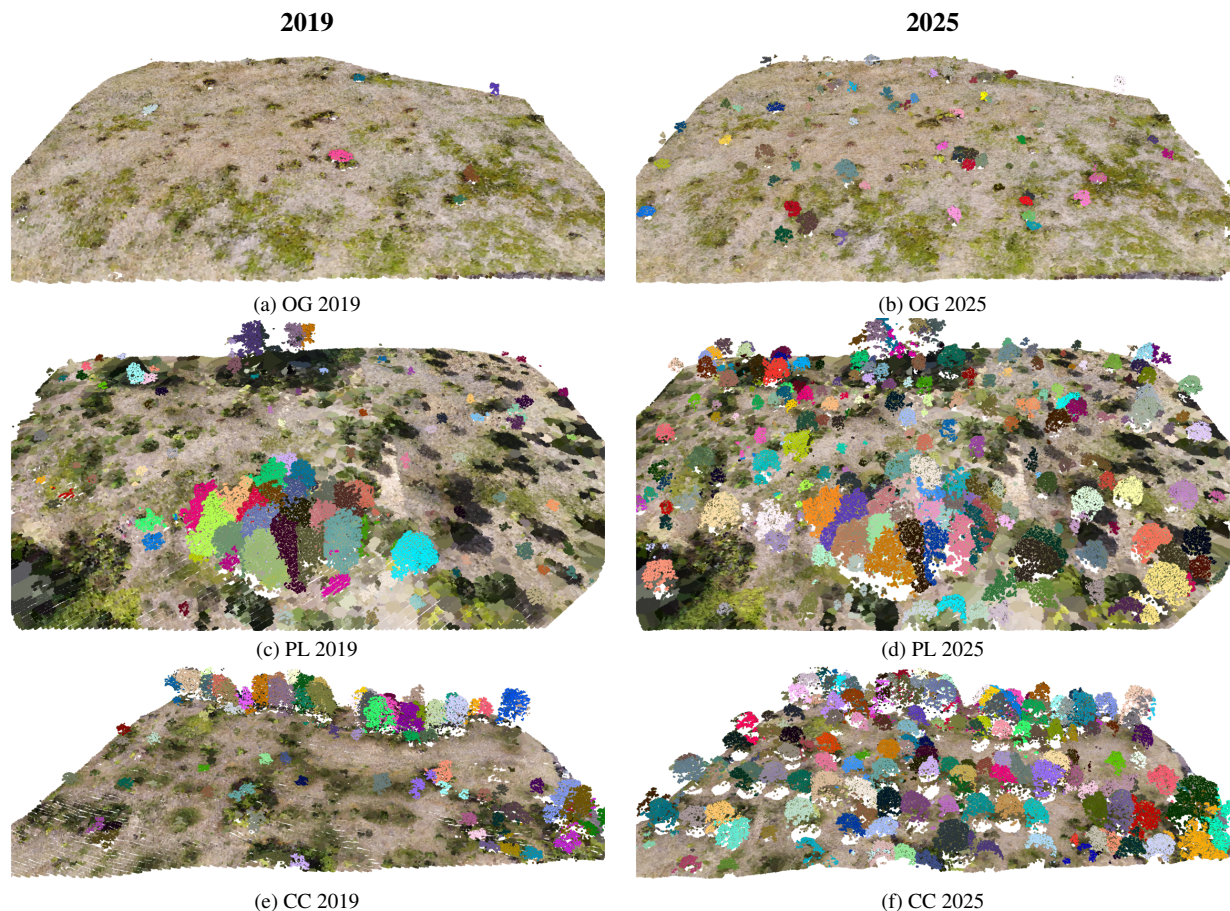


Figure 4. LiDAR segmentation results comparing forest restoration progress across three study areas between 2019 and 2025. Top row: Closed Canopy (CC), Middle row: Park Land (PL), Bottom row: Open Grassland (OG). Left column shows 2019 baseline conditions, right column shows 2025 restoration outcomes. LiDAR points are colored using reconstruction created from the drone surveys.

such as small infrared range (reducing motion sensing sensitivity) and low image quality (producing coarser images which can affect classification and identification) may affect both the data collected and data processing. [6, 17]. We recommend careful consideration of experimental set up and biases (e.g., camera traps pointing at the ground will capture fewer birds) when placing sensors for a study. Additionally, we recommend tests to ascertain camera trap image quality, detection range, detection rates, battery life, and suitability of SD cards such that they match intended use.

Similarly, the limited amount of acoustic sensors available may affect our information at each site and may bias information recorded in Table 4. In addition, bird occurrence can be influenced by various factors such as food availability, seasonality, or the presence of water sources nearby. We also note the presence of more mature patches of forest in the vicinity of PL and CC (Figures 1, 4d and 4f) which may strengthen the dissimilarity of these sites with OG where the native forest is further away from our sensors.

Data processing. Regarding data processing, our measures of richness, occurrence, evenness and Shannon diversity are directly derived from predictions of pre-trained models (BioCLIP for camera traps, and Perch for acoustics). While machine learning models are typically less performant for less represented areas such as Hawai'i [14], we may assume that any systematic prediction errors remain consistent across sites. For example, no cackling goose has ever been observed by field technicians in any of the sites, so cackling geese recorded by PAM are most likely to be nēnēs. However, sites with more cackling geese also have more nēnēs. Hence, our absolute numbers of occurrences may not be reliable, but the difference of diversity measured across sites are more representative of the true relative differences in the field.

Misclassifications also justify the use of soundscape variance to approximate biodiversity as it is robust to potential prediction errors. By combining measures directly derived from species predictions (e.g., species richness, even-

Table 7. **Forest restoration statistics for Closed Canopy (CC), Open Grassland (OG), and Park Land (PL) areas comparing 2019 and 2025 LiDAR data using robust statistics.** We report median and 5%-95% percentile ranges for tree height, diameter, crown area, and crown volume.

Area	Metric	2019			2025		
		Trees	Median	5%-95% Range	Trees	Median	5%-95% Range
OG (9052.6 m ²)	Height (m)	7	1.90	1.55 – 3.94	64	2.68	1.17 – 4.09
	Diameter (m)		2.85	1.99 – 3.96		2.09	1.48 – 3.57
	Crown Area (m ²)		6.39	3.13 – 12.46		3.44	1.73 – 10.03
	Crown Volume (m ³)		9.91	4.76 – 16.31		6.95	1.9 – 20.29
PL (13523.3 m ²)	Height (m)	97	3.12	1.73 – 13.35	216	5.31	2.31 – 13.77
	Diameter (m)		3.69	1.7 – 8.4		4.58	1.8 – 8.72
	Crown Area (m ²)		10.69	2.27 – 55.44		16.50	2.55 – 59.73
	Crown Volume (m ³)		23.50	3.64 – 316.35		56.66	5.02 – 288.86
CC (9698.6 m ²)	Height (m)	78	8.02	2.98 – 13.38	170	8.53	3.18 – 13.1
	Diameter (m)		6.31	3.26 – 9.18		6.76	4.0 – 9.23
	Crown Area (m ²)		31.31	8.33 – 66.27		35.88	12.54 – 66.94
	Crown Volume (m ³)		123.49	19.84 – 381.55		195.32	31.74 – 438.8

ness and Shannon diversity) and soundscape variance, we aim to achieve a reliable estimate of true biodiversity. This automatic processing approach can easily scale up to more sensors, ultimately strengthening the robustness of biodiversity estimate.

Missing information. Finally, to draw ecological conclusions regarding the effects of koa tree restoration, we would need comparisons with both undisturbed native forests and non-restored plots of land. Additionally, records of tree planting events would enable us to link tree planting to the recovering understory and biodiversity, as well as success of tree planting, thus allowing us to properly evaluate the effects of such undertakings.

6. Conclusion

In this study, we propose a methodology that integrates multi-sensory information in order to monitor the variations in biodiversity and forest maturity across three different sites undergoing active koa tree restoration in Hawai'i. We show evidences that as planted areas exhibit higher tree heights and larger crowns, they also exhibit a higher bird species evenness and Shannon diversity. These results support our hypothesis that the ongoing restoration efforts are successful at sustaining the associated biodiversity of these ecosystems, an essential process for their long term stability.

Importantly, our data processing pipeline uses off-the-shelf foundation models to predict species labels from camera trap and passive acoustic monitoring data, along with a simple yet effective unsupervised algorithm to segment in-

dividual trees from LiDAR point clouds. With this fully automatic process, we illustrate how computer vision may be used to drastically scale up the monitoring of forest restoration progress. While we acknowledge the limitations of foundation models and discuss the impact of misclassifications on the derived metrics, we expect that model performance will continue to improve, making our methodology increasingly suitable to larger areas and longer time spans.

7. Acknowledgment

This work was supported by the National Science Foundation OAC 2118240 Imageomics Institute and OISE 2330423 AI and Biodiversity Change Global Center awards and the Natural Sciences and Engineering Research Council of Canada 585136 award. This material draws on research supported in part by the Social Sciences and Humanities Research Council. This material is based in part upon work supported by the National Ecological Observatory Network (NEON), a program sponsored by the U.S. National Science Foundation (NSF) and operated under cooperative agreement by Battelle. We would like to thank the staff of the National Ecological Observatory Network Domain 20 (Pacific Tropical): Shea Uehana, Eissas Ouk, Evan Donoso, Avery Dean, Ann Carey, Maggie Bahr, Daniel Berke, Darien Matlock, Emmy Van Gundy, and Josh Landkamer for their support and efforts in data collection for this study.

References

- [1] Sara Beery, Dan Morris, and Siyu Yang. Efficient pipeline for camera trap image review, 2019. [2](#)
- [2] Marc Besson, Jamie Alison, Kim Bjerger, Thomas E Gorochowski, Toke T Høye, Tommaso Jucker, Hjalte MR Mann, and Christopher F Clements. Towards the fully automated monitoring of ecological communities. *Ecology letters*, 25(12):2753–2775, 2022. [1](#)
- [3] J. Roger Bray and J. T. Curtis. An ordination of the upland forest communities of southern wisconsin. *Ecological Monographs*, 27(4):325–349, 1957. [3](#)
- [4] George C Cummins, Tad C Theimer, and Eben H Paxton. Responses to terrestrial nest predators by endemic and introduced hawaiian birds. *Ecology and Evolution*, 10(4):1949–1958, 2020. [1](#)
- [5] Tom Denton, Scott Wisdom, and John R. Hershey. Improving bird classification with unsupervised sound separation. *arXiv preprint arXiv:2110.03209*, 2021. Accessed: 2025-04-27. [3](#)
- [6] Philip D DeWitt and Amy G Cocksdedge. A simple framework for maximizing camera trap detections using experimental trials. *Environmental Monitoring and Assessment*, 195(11):1381, 2023. [7](#)
- [7] Lucas B Fortini, Adam E Vorsino, Fred A Amidon, Eben H Paxton, and James D Jacobi. Large-scale range collapse of hawaiian forest birds under climate change and the need 21st century conservation options. *PloS one*, 10(10):e0140389, 2015. [1](#)
- [8] E E Z Geneviev, C Price, N Evans, J P Streicher, and C T Downs. Population dynamics and morphometrics of nile monitors along a gradient of urbanization in KwaZulu-Natal, south africa. *J. Zool.* (1987), 326(1):23–36, 2025. [2](#)
- [9] Burooj Ghani, Tom Denton, Stefan Kahl, and Holger Klinck. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports*, 13(1):22876, 2023. [3](#)
- [10] Jacob Goldberger, Geoffrey E Hinton, Sam Roweis, and Russ R Salakhutdinov. Neighbourhood components analysis. *Advances in neural information processing systems*, 17, 2004. [4](#)
- [11] Andres Hernandez, Zhongqi Miao, Luisa Vargas, Sara Beery, Rahul Dodhia, and Juan Lavista. Pytorch-wildlife: A collaborative deep learning framework for conservation, 2024. [2](#)
- [12] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970. [4](#)
- [13] Paul Jaccard. Étude comparative de la distribution florale dans une portion des alpes et des jura. *Bulletin de la Société vaudoise des sciences naturelles*, 37(142):547–579, 1901. [3](#)
- [14] Stefan Kahl, Amanda Navine, Tom Denton, Holger Klinck, Patrick Hart, Hervé Glotin, Hervé Goëau, Willem-Pier Vellinga, Robert Planqué, and Alexis Joly. Overview of birdclef 2022: Endangered bird species recognition in soundscape recordings. CEUR-WS, 2022. [7](#)
- [15] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International journal of computer vision*, 1, 1988. [3](#)
- [16] Morgane Labadie, Serge Morand, Mathieu Bourgarel, Fabien Roch Niama, Guytrich Franel Ngulili, N’kaya Tobi, Alexandre Caron, and Helene De Nys. Habitat sharing and interspecies interactions in caves used by bats in the republic of congo. *PeerJ*, 13:e18145, 2025. [2](#)
- [17] J Marcus Rowcliffe, Chris Carbone, Patrick A Jansen, Roland Kays, and Bart Kranstauber. Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, 2(5):464–476, 2011. [7](#)
- [18] Jari Oksanen, Gavin L. Simpson, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R.B. O’Hara, Peter Solymos, M. Henry H. Stevens, Eduard Szoecs, Helene Wagner, Matt Barbour, Michael Bedward, Ben Bolker, Daniel Borcard, Tuomas Borman, Gustavo Carvalho, Michael Chirico, Miquel De Caceres, Sebastien Durand, Heloisa Beatriz Antoniazzi Evangelista, Rich FitzJohn, Michael Friendly, Brendan Furneaux, Geoffrey Hannigan, Mark O. Hill, Leo Lahti, Cameron Martino, Dan McGlinn, Marie-Helene Ouellette, Eduardo Ribeiro Cunha, Tyler Smith, Adrian Stier, Cajo J.F. Ter Braak, and James Weedon. *vegan: Community Ecology Package*, 2025. R package version 2.8-0. [3](#)
- [19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011. [4](#)
- [20] Angela J. L. Pestell, Anthony R. Rendall, Robin D. Sinclair, Euan G. Ritchie, Duc T. Nguyen, Dean M. Corva, Anne C. Eichholtzer, Abbas Z. Kouzani, and Don A. Driscoll. Smart camera traps and computer vision improve detections of small fauna. *Ecosphere*, 16(3):e70220, 2025. [2](#)
- [21] E C Pielou. The measurement of diversity in different types of biological collections. *J. Theor. Biol.*, 13:131–144, 1966. [3](#)
- [22] Laura J Pollock, Justin Kitzes, Sara Beery, Kaitlyn M Gaynor, Marta A Jarzyna, Oisín Mac Aodha, Bernd Meyer, David Rolnick, Graham W Taylor, Devis Tuia, et al. Harnessing artificial intelligence to fill global shortfalls in biodiversity knowledge. *Nature Reviews Biodiversity*, pages 1–17, 2025. [1](#)
- [23] Hugh P. Possingham, Sandy J. Andelman, Barry R. Noon, Stephen Trombulak, and Howard R. Pulliam. Making smart conservation decisions. In *Conservation Biology: Research Priorities for the Next Decade*, pages 229–244. Island Press, Washington, D.C., 2001. Accessed: 2025-04-25. [1](#)
- [24] Joris Ravaglia, Alexandra Bac, and Richard A. Fournier. Extraction of tubular shapes from dense point clouds and application to tree reconstruction from laser scanned data. *Computers & Graphics*, 66:23–33, 2017. Shape Modeling International 2017. [3](#)
- [25] Michael D Samuel, Peter HF Hobbelen, Francisco DeCastro, Jorge A Ahumada, Dennis A LaPointe, Carter T Atkinson, Bethany L Woodworth, Patrick J Hart, and David C Duffy. The dynamics, transmission, and population impacts of avian

malaria in native hawaiian birds: a modeling approach. *Ecological Applications*, 21(8):2960–2973, 2011. 1

- [26] Paul G Scowcroft and Jack Jeffrey. Potential significance of frost, topographic relief, and acacia koa stands to restoration of mesic hawaiian forests on abandoned rangeland. *Forest Ecology and Management*, 114(2-3):447–458, 1999. 1
- [27] C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. 3
- [28] Ashley T. Simkins, William J. Sutherland, Lynn V. Dicks, Craig Hilton-Taylor, Molly K. Grace, Stuart H. M. Butchart, Rebecca A. Senior, and Silviu O. Petrovan. Past conservation efforts reveal which actions lead to positive outcomes for species. *PLOS Biology*, 23:e3003051, 2025. Accessed: 2025-04-25. 1
- [29] Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, et al. Bioclip: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 19412–19424, 2024. 2
- [30] Laurie Strommer and Sheila Conant. Conservation value of koa (acacia koa) reforestation areas on hawaii island. *Pacific Conservation Biology*, 24(1):35–43, 2018. 2
- [31] Yi-Hsing Tseng, Kai-Pei Tang, and Fu-Chen Chou. Surface reconstruction from lidar data with extended snake theory. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 479–492, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. 3
- [32] Charles van Riper III, J Michael Scott, et al. Limiting factors affecting hawaiian native birds. *Studies in Avian Biology*, 22(1):28, 2001. 1
- [33] Scott Wisdom, Efthymios Tzinis, Hakan Erdogan, Ron Weiss, Kevin Wilson, and John Hershey. Unsupervised sound separation using mixture invariant training. *Advances in neural information processing systems*, 33:3846–3857, 2020. 3
- [34] Juntao Yang, Zhizhong Kang, Sai Cheng, Zhou Yang, and Perpetual Akwensi. An individual tree segmentation method based on watershed algorithm and 3d spatial distribution analysis from airborne lidar point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13:1055–1067, 2020. 3
- [35] Kaiguang Zhao and Sorin Popescu. Hierarchical watershed segmentation of canopy height model for multi-scale forest inventory. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(3/W52): 436–441, 2007. 3

A. Species list

Common Name	Scientific Name	Status
'Akiapōlā'au	<i>Hemignathus munroi</i>	Native
'Apapane	<i>Himatione sanguinea</i>	Native
Cackling Goose	<i>Branta hutchinsii</i>	Non-Native
Common Waxbill	<i>Estrilda astrild</i>	Non-Native
Hawai'i 'amakihi	<i>Chlorodrepanis virens</i>	Native
Hawai'i creeper	<i>Loxops mana</i>	Native
Nēnē	<i>Branta sandvicensis</i>	Native
Japanese Bush Warbler	<i>Horornis diphone</i>	Non-Native
Northern Cardinal	<i>Cardinalis cardinalis</i>	Non-Native
'Ōma'o	<i>Myadestes obscurus</i>	Native
Pueo*	<i>Asio flammeus sandwichensis</i>	Native
Spotted Dove	<i>Spilopelia chinensis</i>	Non-Native
Warbling White-eye	<i>Zosterops japonicus</i>	Non-Native
Wild Turkey	<i>Meleagris gallopavo</i>	Non-Native
Yellow-fronted Canary	<i>Serinus mozambicus</i>	Non-Native
Kalij pheasant	<i>Lophura leucomelanos</i>	Non-Native
American barn owl	<i>Tyto furcata</i>	Non-Native

Table 8. **Common and scientific names of bird species in camera trap or bioacoustic data.** *Pueo is a subspecies of the short-eared owl.

B. iNaturalist Dataset

We downloaded images of the 7 bird species and 5 plant species: *Acacia koa* (Koa), *Metrosideros polymorpha* ('Ōhi'a), *Coprosma ernodeoides* (Kūkaenēnē), *Coprosma montana* (Pilo), and *Cibotium glaucum* (Hāpu'u) from iNaturalist². We ran MegaDetector on these images and obtained 1330 'animal' crops.

C. Annotated data set

Category	Total	Train	Test
Nēnē	832	582	250
Hawai'i 'amakihi	3	2	1
Common waxbill	2	1	1
'Apapane	5	4	1
Kalij pheasant	28	20	8
'Ōma'o	9	6	3
American barn owl	2	1	1
Non-bird	2477	1734	743

Table 9. **Train test split of the annotated camera trap dataset.**

²<https://www.inaturalist.org/>