Towards Anomaly Detection in Forest Biodiversity Monitoring: A Pilot Study with Variational Autoencoders

David Susič
david.susic@ijs.si
Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

Maria Luisa Buchaillot
Fauna Smart Technologies ApS
Copenhagen, Denmark

Miguel Crozzoli
Intelligent Instruments Lab,
University of Iceland
Reykjavik, Iceland

Calum Builder
Fauna Smart Technologies ApS
Copenhagen, Denmark

Sevasti Maistrou
Fauna Smart Technologies ApS
Copenhagen, Denmark

Anton Gradišek
Department of Intelligent Systems,
Jožef Stefan Institute
Ljubljana, Slovenia

Dragana Vukašinović
Fauna Smart Technologies ApS
Copenhagen, Denmark

Abstract

Biodiversity monitoring in forests requires scalable, automated tools for detecting ecological anomalies across time and space. This paper reports on a three-month pilot deployment (April 1 to June 30, 2025) in Dyrehaven, an 11 km^2 forest park near Copenhagen, Denmark, where acoustic data from 10 distributed AudioMoth sensors and vegetation indices from Sentinel-2 imagery were collected. We trained separate variational autoencoder (VAE) models on each modality to test the technical feasibility of learning ecological baselines. Since no ecological anomalies occurred during the observation period, evaluation focused on reconstruction errors, which indicate how well VAEs can capture typical site-specific ecological patterns (i.e., baseline modeling). Both acoustic and satellite pipelines achieved low reconstruction errors, demonstrating that VAEs can reliably model normal ecological dynamics. This establishes the foundation for future studies on anomaly detection, which will require larger datasets containing true ecological anomalies identified and labeled by experts. Ongoing work focuses on extending data collection to additional forest sites, while future anomaly detection will require expert-labeled anomalies to calibrate baselines and validate model performance for robust, multimodal biodiversity monitoring.

Keywords

biodiversity, anomaly detection, variational autoencoder, machine learning, passive acoustic monitoring, satellite imagery

1 Introduction

Forests are complex, dynamic ecosystems increasingly affected by environmental stressors such as pests, diseases, invasive species, and climate-related disturbances [1]. Effective biodiversity monitoring is essential to detect these stressors early and support adaptive, science-based forest management [2, 3]. However, existing monitoring tools are often limited in scope, fragmented across disciplines, and costly to implement at scale [4].

This paper presents the technical foundation of the biodiversity assessment tool (BAT), a modular, scalable system that integrates ecoacoustics, satellite remote sensing, and machine learning (ML) to enable automated biodiversity monitoring in forested landscapes. BAT is designed to detect anomalies in ecological baselines, providing early warning signals of ecosystem degradation [5]. It combines two complementary remote sensing modalities: passive acoustic monitoring (PAM), which captures localized, high-frequency biological activity such as insect or bird calls [6, 7], and satellite Earth observation (EO), which offers broader, lower-frequency indicators of landscape-level change, including vegetation health and canopy dynamics [8].

The presence of pests or other stressors often leads to a reduction in biodiversity, which can first be detected acoustically as diminished biotic sound activity, and later (typically with a lag of several days) becomes visible in EO data as decreased vegetation greenness. BAT is designed to leverage this temporal and spatial complementarity by developing independent anomaly detection pipelines for each modality, which in future iterations may support joint multimodal detection of ecological disturbances.

This study reports on a pilot deployment in Dyrehaven, a human-managed park-forest in Denmark, where time-series data from distributed acoustic sensors and Sentinel-2 satellite imagery were collected between April and June 2025. Separate variational autoencoders (VAEs) were trained on each modality to test whether robust baseline models can be learned. Ecological anomalies are inherently rare and cannot be guaranteed within a limited three-month window, and none occurred during this period. As a result, evaluation focused on baseline reconstruction performance rather than anomaly detection accuracy. Demonstrating that VAEs can successfully capture "normal" ecological patterns is a necessary prerequisite for future anomaly detection. Ecological baselines are inherently site-specific, differing across forest types, microhabitats, and even within single forests (e.g., wetter zones near ponds vs drier uplands). Accordingly, this work should be understood as a technical feasibility study, with the longer-term goal of enabling multimodal detection of ecological disturbances such as pest outbreaks, supported by expert-labeled events and extended deployment across diverse forests.

Information Society 2025, Ljubljana, Slovenia

https://doi.org/10.70314/is.2025.skui.6757

2 Data

Our study area was Dyrehaven, a human-managed forest park north of Copenhagen, Denmark (55.8024°N, 12.5685°E), covering 11 km^2 (see Figure 1). The site includes 10 structured microhabitats across woodland, meadow, and modified forest areas. Its ecological diversity and relative stability make it suitable for testing acoustic and satellite-based monitoring methods. Data were collected between April 1 and June 30, 2025.

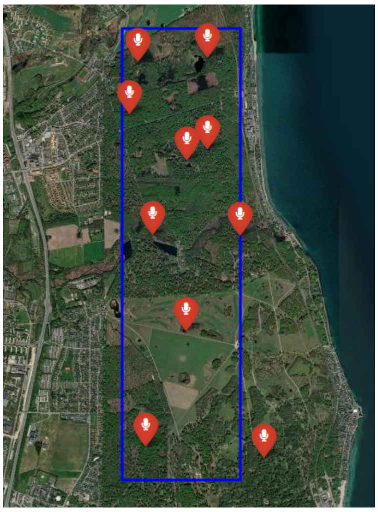

Figure 1: Study area in Dyrehaven, Denmark with AudioMoth recording locations (red pins) and Sentinel-2 satellite bounding box (blue).

2.1 Audio

Passive acoustic data were collected using 10 AudioMoth recording devices deployed across Dyrehaven's microhabitats. Devices were positioned to maximize spatial heterogeneity, minimize acoustic overlap, and ensure temporal consistency. Each unit recorded 45-second mono-channel clips every five minutes at a 48 kHz sampling rate. All devices were weatherproofed and mounted on trees for continuous outdoor operation. A recording gap occurred between April 20 and April 29 due to memory card failure. A total of 203078 recordings were generated during the study period. After removing corrupted or incomplete files (309 clips, 0.15%), 202769 valid recordings remained.

2.2 Visual

Satellite imagery was sourced from the Sentinel-2 mission [9], covering a 1.48 km × 5.86 km bounding box encompassing 9 of the 10 AudioMoth locations. Out of 53 total available snapshots during our study period, 18 cloud-free scenes (≤50% cloud cover) were selected for analysis to ensure index reliability.

Normalized difference vegetation index (NDVI) and Normalized difference moisture index (NDMI) were computed for each selected image as

$$\text{NDVI} = \frac{\text{NIR} - \text{red}}{\text{NIR} + \text{red}}$$

and

$$\text{NDMI} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}},$$

where, NIR, SWIR, and red are near-infrared, shortwave-infrared, and visible red bands, respectively.

NDVI was calculated at 10 m resolution, and NDMI at 20 m. Each index map was divided into fixed-size patches. NDVI maps produced 396 patches (11 × 36 grid), while NDMI produced 108 patches (6×18 grid), reflecting their respective spatial resolutions.

3 Methodology

3.1 Extraction of Acoustic Indices

10 standard ecoacoustic indices [10] (list in Table 1) were extracted from each 45-second recording, capturing patterns from both time-domain and time-frequency analyses. These indices reflect aspects such as spectral entropy, acoustic complexity, temporal dynamics, and frequency distribution, offering proxies for ecological features like species richness, biophonic activity, and anthropogenic disturbance. All indices were independently normalized to the [0, 1] range using their dataset-wide minimum and maximum values.

Table 1: Acoustic indices used in this study and their ecological interpretation.

Index	Use
ACI	Detects dynamic biotic sounds (e.g., bird choruses).
AEI	Identifies dominance vs. diversity in acoustic communities.
EAS	Differentiates uniform noise vs. structured signals.
ECU	Indicates unpredictability and complexity of soundscapes.
ECV	Captures temporal structure (e.g., insect or bird rhythms).
EPS	Distinguishes tonal vs. noisy sound environments.
ADI	Proxy for acoustic diversity or species richness.
NDSI	Separates natural from human-made noise.
Ht	Detects continuous vs. discrete acoustic events.
ARI	Estimates overall acoustic richness.

3.2 Preprocessing of Satellite Imagery

To ensure patch-level data quality, we applied the scene classification layer (SCL) after resampling. Patches containing cloudy or unreliable pixels (SCL classes 3, 8, 9, or 10) were excluded. This preprocessing pipeline produced curated spatiotemporal datasets of 4436 NDVI patches and 1226 NDMI patches, which served as input for training and evaluating the VAE models.

3.3 Variational Autoencoder and Evaluation Metrics

A variational autoencoder (VAE) learns to compress input data into a latent representation and reconstruct it via encoder and decoder as per Figure 2.

The encoder maps each input to a latent mean, μ_1 and log-variance, $log(\sigma_1^2)$, from which a latent vector z is sampled via the reparameterization trick: $z = \mu_1 + \sigma_1 \cdot \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ and $\sigma_1 = \exp(0.5 \cdot \log(\sigma_1^2))$.

The decoder reconstructs the input from z, producing a mean μ_2 and log-variance $\log(\sigma_2^2)$ of the output distribution. Training minimizes the total loss:

$$\mathcal{L}_{\text{VAE}} = \mathcal{L}_{\text{recon}} + w_{\text{KL}} \cdot \mathcal{L}_{\text{KL}}$$

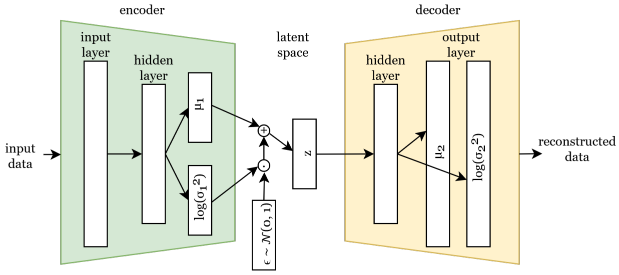

Figure 2: Architecture of VAE for anomaly detection using reconstruction probability.

where $\mathcal{L}_{\text{recon}}$ is the negative log-likelihood of the input under the decoder's Gaussian output:

$$\mathcal{L}_{\text{recon}} = -\sum_{i=1}^{D} \log \mathcal{N}(x_i \mid \mu_{2,i}, \sigma_{2,i}^2)$$

and $\mathcal{L}_{\text{KL}}$ is the Kullback–Leibler divergence between the approximate posterior $q(z|x)$ and the prior $p(z) = \mathcal{N}(0,1)$:

$$\mathcal{L}_{\text{KL}} = -\frac{1}{2}\sum_{j=1}^{d}\left(1 + \log(\sigma_{1,j}^2) - \mu_{1,j}^2 - \sigma_{1,j}^2\right)$$

with D and d representing the input and latent dimensions, respectively.

In an operational anomaly detection setting, the decoder's negative log-likelihood (often referred to as reconstruction likelihood) would serve as the anomaly score, with higher values indicating more anomalous inputs. However, since no ecological anomalies occurred during our three-month observation window, this pilot study evaluates baseline modeling rather than anomaly detection accuracy. Specifically, we report reconstruction errors: mean squared error (MSE) and mean absolute error (MAE) for acoustic indices, and overall mean absolute error (averaged across all pixels in each patch) for NDVI and NDMI patches, computed only on non-cloudy patches after SCL masking.

3.4 Experimental Setup

The general pipeline of the BAT system is shown in Figure 3. It consists of independent audio and visual pipelines designed to operate separately but eventually integrate into a unified decision-support framework.

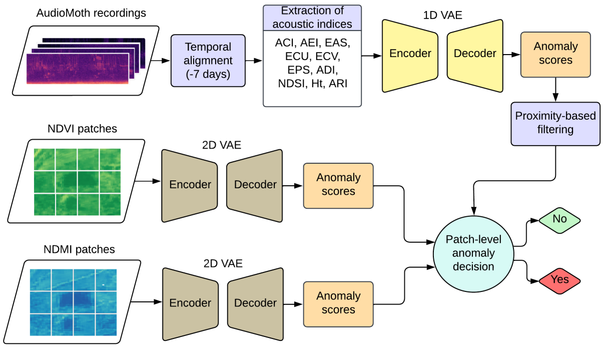
AudioMoth recordings
Temporal alignment (-7 days)
Extraction of acoustic indices
ACI, AEI, EAS, ECU, ECV, EPS, ADI, NDSI, Ht, ARI
1D VAE
Encoder
Decoder
Anomaly scores
Proximity-based filtering
NDVI patches
2D VAE
Encoder
Decoder
Anomaly scores
NDMI patches
2D VAE
Encoder
Decoder
Anomaly scores
Patch-level anomaly decision
No
Yes

Figure 3: The general pipeline of the BAT system.

In a full anomaly detection setting, the pipelines would use reconstruction likelihoods as anomaly scores and combine them across modalities. In this pilot, since no anomalies occurred, we only assess baseline modeling by training and evaluating the acoustic and satellite VAEs independently, reporting reconstruction errors as indicators of model performance.

3.4.1 *Audio Pipeline.*

The audio VAE uses a 10-dimensional input, with an encoder and decoder each containing one hidden layer of size 8 and ReLU activation. The latent space has dimension 4. The decoder outputs the reconstructed mean and log-variance of size 10.

Model evaluation used 5-fold cross-validation with folds defined by spatially clustered AudioMoth devices (~ 850 m minimum separation) to reduce data leakage. Models were trained for 30 epochs with a batch size of 512 using the Adam optimizer and a one-cycle learning rate schedule.

3.4.2 *Visual Pipeline.*

The satellite VAE takes a 16×16 pixel input (NDVI or NDMI) and uses three convolutional layers (32, 64, 128 filters) with ReLU activation in the encoder. The output is flattened and mapped to a latent space of dimension 4. The decoder upsamples using three transposed convolutional layers with ReLU, reconstructing the mean and log-variance patches of size 16×16.

Separate VAE models were trained for NDVI and NDMI using an 80/20 train-test split. Each model was trained for 20 epochs with a batch size of 32 using the Adam optimizer. The loss was computed only over non-cloudy pixels.

4 Results and Discussion

To examine temporal patterns, all indices were plotted over the study period as seen in Figure 4. Acoustic indices were averaged between 9AM and 3PM across all 10 AudioMoth devices to avoid nighttime inactivity and minimize dawn/dusk transitions. A 10-day smoothing window was applied to reduce day/night fluctuations. The indices remained relatively stable long-term, showing little trend and suggesting no major ecological disruptions and reflecting the stability of the forest soundscape over the study period.

Visual indices were averaged across all patches for each date. Both indices exhibit a gradual increase from early April to late June, consistent with seasonal greening. NDVI shows a smooth and consistent rise, indicating widespread vegetation growth. NDMI, while generally increasing, displays more irregular variation, particularly early in the season, likely reflecting transient moisture conditions. NDVI primarily tracks canopy structure and greenness, while NDMI is more sensitive to vegetation and soil moisture.

The audio pipeline VAE was evaluated using reconstruction MSE and MAE. Since all indices were normalized to the [0,1] range, errors are directly comparable. As shown in Figure 5, reconstruction errors are generally low, indicating that the model effectively captures the underlying structure of the acoustic data.

EPS and Ht showed the highest reconstruction error variability. This suggests they are more difficult to model but may provide sensitive signals of ecological change in future anomaly detection settings. Indices with consistently low reconstruction errors, on the other hand, indicate stable features that can serve as robust components of ecological baselines. These patterns highlight differences in how well various indices represent typical acoustic dynamics, which is central to establishing reliable baseline models.

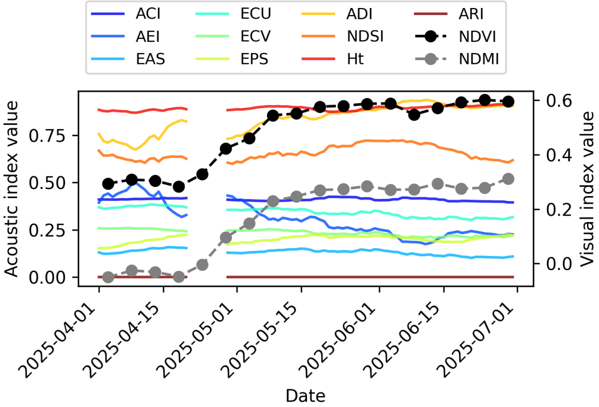

Figure 4: Index values over the study period.

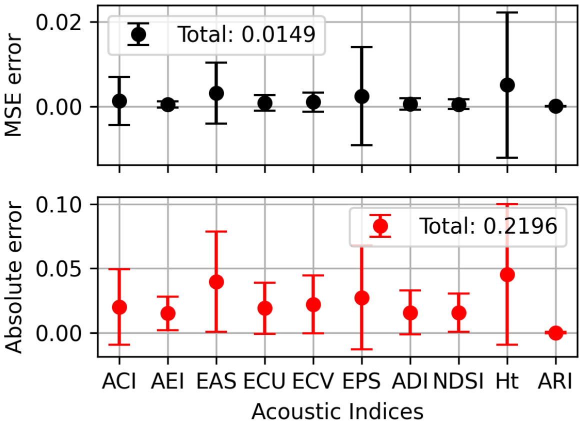

Figure 5: Reconstruction errors for acoustic indices.

The visual pipeline VAEs were evaluated using overall MAE per patch. As expected, errors were fairly uniform across pixels, indicating that the models reconstruct spatial patterns consistently without localized distortions. The average patch-level MAE (average across all $16 \times 16 = 256$ pixels across all images) was 7.17 ± 0.11 for NDVI and 9.65 ± 0.26 for NDMI. Given the [0, 1] normalization range of each pixel, the errors are relatively small and therefore reflect accurate reconstruction of vegetation and moisture dynamics.

The selected VAE models for both the acoustic and visual pipelines demonstrate strong reconstruction performance, with consistently low errors across acoustic indices and Sentinel-derived NDVI/NDMI patches. This confirms that the models effectively capture typical ecological patterns, which is the intended outcome of this pilot study. While further hyperparameter tuning could potentially reduce errors, the key result is that robust ecological baselines can be modeled. Anomaly detection itself will require expert-labeled events in future deployments, but these results provide the necessary technical foundation.

5 Conclusion

This work demonstrates the technical feasibility of using VAEs to model baseline ecological patterns from acoustic and satellite time series in a forested landscape. As a pilot study, it does not evaluate anomaly detection directly, since no anomalies occurred during the observation period. Instead, it establishes that robust models can be trained on available data, providing a foundation for future multimodal monitoring.

A critical next step is the collection of additional data over longer time frames and across multiple forest types, since actual ecological anomalies are rare and cannot be guaranteed within a short observation window. Detecting and validating anomalies will require expert labeling of such events once they occur. To this end, we are continuing data collection at Dyrehaven and planning expansions to other Danish forests (e.g., Thy, Amager, Lillebælt) to capture a wider range of ecological contexts and improve model generalization. Further development will also focus on refining acoustic preprocessing through time-window averaging or time-aware features and enhancing the visual pipeline with seasonal baselines, sequential models, and zone-specific approaches that account for spatial heterogeneity.

With expert input, longer-term recordings, and broader deployment, the BAT system can evolve from modeling site-specific baselines into a robust anomaly detection tool supporting scalable and long-term biodiversity monitoring.

Acknowledgements

This work was funded by Fauna Smart Technologies ApS under the European Space Agency (ESA) grant no. 4000147116, Biodiversity Assessment Tool.

References

[1] William R. L. Anderegg, Oriana S. Chegwidden, Grayson Badgley, Anna T. Trugman, Danny Cullenward, John T. Abatzoglou, Jeffrey A. Hicke, Jeremy Freeman, and Joseph J. Hamman. 2022. Future climate risks from stress, insects and fire across us forests. *Ecology Letters*, 25, 6, 1510–1520. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.14018. DOI: https://doi.org/10.1111/ele.14018.

[2] Lucas P. Gaspar et al. 2023. Predicting bird diversity through acoustic indices within the atlantic forest biodiversity hotspot. *Frontiers in Remote Sensing*, 4, (Dec. 2023). DOI: 10.3389/frsen.2023.1283719.

[3] J.Wolfgang Wägele et al. 2022. Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology*, 59, 105–138. DOI: https://doi.org/10.1016/j.baae.2022.01.003.

[4] Santiago Izquierdo-Tort, Andrea Alatorre, Paulina Arroyo-Gerala, Elizabeth Shapiro-Garza, Julia Naime, and Jérôme Dupras. 2024. Exploring local perceptions and drivers of engagement in biodiversity monitoring among participants in payments for ecosystem services schemes in southeastern mexico. *Conservation Biology*, 38, 6, e14282. eprint: https://conbio.onlinelibrary.wiley.com/doi/pdf/10.1111/cobi.14282. DOI: https://doi.org/10.1111/cobi.14282.

[5] Nathalie Pettorelli, Jake Williams, Henrike Schulte to Bühne, and Merry Crowson. 2025. Deep learning and satellite remote sensing for biodiversity monitoring and conservation. *Remote Sensing in Ecology and Conservation*, 11, 2, 123–132. eprint: https://zslpublications.onlinelibrary.wiley.com/doi/pdf/10.1002/rse2.415. DOI: https://doi.org/10.1002/rse2.415.

[6] Rory Gibb, Ella Browning, Paul Glover-Kapfer, and Kate E. Jones. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10, 2, 169–185. eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.13101. DOI: https://doi.org/10.1111/2041-210X.13101.

[7] D.A. Nieto-Mora, Susana Rodríguez-Buritica, Paula Rodríguez-Marín, J.D. Martínez-Vargaz, and Claudia Isaza-Narváez. 2023. Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. *Heliyon*, 9, 10, e20275. DOI: https://doi.org/10.1016/j.heliyon.2023.e20275.

[8] Nathalie Pettorelli et al. 2018. Satellite remote sensing of ecosystem functions: opportunities, challenges and way forward. *Remote Sensing in Ecology and Conservation*, 4, 2, 71–93. eprint: https://zslpublications.onlinelibrary.wiley.com/doi/pdf/10.1002/rse2.59. DOI: https://doi.org/10.1002/rse2.59.

[9] Copernicus Data Space Ecosystem. 2015. Sentinel-2. (2015). https://dataspace.copernicus.eu/explore-data/data-collections/sentinel-data/sentinel-2.

[10] Luis J. Villanueva-Rivera, Bryan C. Pijanowski, Jarrod Doucette, and Burak Pekin. 2011. A primer of acoustic analysis for landscape ecologists. *Landscape Ecology*, 26, 9, (July 2011), 1233–1246. DOI: 10.1007/s10980-011-9636-9.