Project plan for degree project

Template version 1.6 – June 12, 2020

DV2572: MASTER'S THESIS (120 CREDITS) IN COMPUTER SCIENCE

February 8, 2026

Title	Data-Centric AI for Acoustic Discovery: Anomaly Detection and Active Learning in Bioacoustics.	
Classification	Machine Learning, Anomaly detection, Bioacoustics	
Student 1	Name	Urranki Kamaleswar
	E-Mail	kaur24@student.bth.se
	Person nr	20040106-T197
	Program	Master's Programme in Computer Science
Student 2	Name	Vineeth Dhagey
	E-Mail	vidh24@student.bth.se
	Person nr	20030821-T176
	Program	Master's Programme in Computer Science
Supervisor	Name & title	Huseyin Kusetogullari
	E-Mail	huseyin.kusetogullari@bth.se
	Department	Department of Computer Science and Engineering, BTH
Co-advisor	Name & title	Olof Mogren (Senior Researcher); John Martinsson (Machine learning researcher)
	E-Mail	olof.mogren@ri.se john.martinsson@ri.se
	Company	RISE Research Institutes of Sweden

*2012 ACM Computing Classification System: www.acm.org/about/class/2012
**Co-advisor from industry or a higher education institution (HEI).

1 Introduction

Long-term ecological monitoring plays a crucial role in understanding biodiversity loss, ecosystem change, and the impacts of climate change. One of the most effective tools for large-scale wildlife observation is passive acoustic monitoring, where continuous sound recordings are collected from natural environments to capture animal vocalisations and broader environmental activity. Advances in sensing technology and data storage now enable ecological observatories to gather thousands of hours of audio data, creating new opportunities to study animal behaviour at scale.

Seabird colonies represent particularly rich and dynamic acoustic environments. The **Stora Karlsö Auk Lab** operates one of the most extensive long-term monitoring programs for the **common guillemot (*Uria aalge*)** in the Baltic Sea. It's continuous multi-channel recordings capture a complex soundscape that includes breeding calls, social interactions, environmental noise, and unusual vocal behaviours. These rare events are scientifically valuable, as they may reveal behavioural changes, stress responses, or previously undocumented communication patterns.

However, the rapid growth of acoustic data has created a significant analysis bottleneck. Although recordings span thousands of hours, only a small fraction contains biologically meaningful events. Manual inspection by experts is time-consuming, costly, and difficult to scale. This challenge is especially severe in discovery-driven bioacoustics, where the objective is not only to recognise known calls but also to uncover rare and potentially novel vocalisations hidden within continuous soundscapes.

Recent advances in artificial intelligence offer promising tools to address this problem. **Unsupervised anomaly detection** can model common background sounds and highlight unusual acoustic patterns without requiring labeled data. **Active learning** introduces a system in which it selects the most informative samples for expert review, thereby reducing labeling effort. In addition, self-supervised learning enables models to learn useful audio representations from large unlabeled datasets, supporting downstream discovery tasks.

Despite these developments, much of the existing research remains benchmark-oriented. Many methods are evaluated on curated datasets with segmented clips, predefined classes, and balanced labels, conditions that rarely reflect real ecological monitoring scenarios. In practice, soundscapes are continuous, annotations are scarce, and rare biological events are diverse and subtle. As a result, anomaly detection systems may highlight sensor noise or environmental artifacts rather than meaningful vocalisations, and active learning strategies often struggle in open-ended discovery settings.

This thesis therefore focuses on developing a **data-centric AI framework for acoustic discovery** that efficiently guides expert attention toward rare and unusual guillemot vocalisations within large unlabeled recordings. By integrating anomaly detection with active learning, the proposed framework aims to maximise discovery efficiency, reduce redundant labeling, and support open-set exploration of previously unseen vocalisation types.

From an ecological perspective, this work contributes to a deeper understanding of seabird behaviour and colony dynamics. From a computational perspective, it advances machine learning toward more realistic, human-efficient, and discovery-driven workflows. Ultimately, the goal is to enable scalable bioacoustic monitoring systems in which expert time is treated as a critical resource and rare ecological phenomena can be identified more effectively.

1.1 Key terminology

This subsection defines key terms used in the thesis to ensure consistent interpretation.

Guillemots: Seabirds in the auk family, commonly referring to species in the genus *Uria*. In this thesis context, the term is used to denote guillemot-related acoustic activity (e.g., vocalizations or sound events) that may appear in marine acoustic recordings.

Bioacoustics: The study and analysis of biological sounds to understand organisms and their interactions with the environment. In marine settings, bioacoustics leverages underwater recordings to detect, characterize, or monitor animal presence and behavior.

Data-Centric AI: A machine learning approach that prioritizes improving the quality, coverage, and consistency of data (e.g., labeling protocols, noise handling, representativeness) as a primary way to improve system performance, rather than focusing only on changing model architectures.

Anomaly Detection: Methods for identifying data segments that deviate from normal or common patterns. In bioacoustics, anomalies can correspond to rare biological sound events, unusual behaviors, or non-biological artifacts such as vessel noise or sensor interference.

Active Learning: A paradigm where the model selectively chooses the most informative unlabeled samples to be labeled by an expert. The goal is to reduce labeling effort while improving performance, especially when labeled data is scarce.

1.2 Related Work

1.2.1 Representation Learning and Embeddings

Modern bioacoustic systems typically begin by transforming raw audio recordings into compact feature representations, often referred to as embeddings. In bird acoustics, pretrained models such as BirdNET provide strong and reusable representations that perform well across a wide range of bird species and recording conditions [22, 23]. More recently, foundation-style bioacoustic models, such as Perch, have been developed to learn general-purpose audio representations that can transfer effectively across tasks and environments [6]. Together, these advances indicate that reliable feature extraction is becoming a mature and practical component of bioacoustic systems. In the context of this thesis, embeddings are particularly valuable because they enable grouping of similar sounds, efficient search within large audio archives, and discovery-oriented analysis without requiring densely labeled datasets from the start[13].

1.2.2 Active Learning and Weak-to-Strong Labels

A major challenge in long-term acoustic monitoring is the amount of time required from domain experts to review and annotate recordings. Active learning (AL) is a widely studied approach for reducing annotation effort by prioritizing the most informative audio segments for labeling instead of selecting samples at random [8, 11]. Common AL strategies include selecting uncertain examples, promoting diversity in the selected samples, or combining both approaches to balance exploration and refinement [9, 18].

At the same time, weak-to-strong label refinement reflects the practical reality that experts often begin with coarse annotations, such as approximate time intervals—and only later refine precise event boundaries. Swedish research on adaptive change-point detection combined with active learning offers a practical method for refining these boundaries using model feedback, improving label quality while keeping annotation effort manageable [10]. Taken together, these studies demonstrate that expert workload can be reduced significantly. However, most existing work primarily focuses on improving detection of known classes or minimizing the number of required labels, rather than explicitly supporting the discovery of novel vocalizations.

1.2.3 Discovery and Practical Workflows

Discovery in settings with limited labeled data is commonly approached through clustering, novelty-aware modelling, and open-set recognition techniques. Studies in ecoacoustics show that clustering and visualization can uncover meaningful structure in long-duration recordings, allowing researchers to explore soundscapes beyond predefined categories [16, 21]. In marine bioacoustics, deep embedded clustering has been used to separate recurring acoustic patterns in complex natural environments, demonstrating that unsupervised discovery can remain effective even under substantial environmental noise [14, 15]. Open-set and novelty detection methods further reinforce the idea that real-world recordings naturally contain sounds that fall outside any initial label set [4, 5, 12].

At a practical level, tools such as OpenSoundscape, Whombat, and LEAVES support reproducible analysis and machine learning–assisted annotation workflows [2, 3, 20]. However, these tools and methodological advances are often developed separately and are rarely integrated into a unified discovery-focused workflow that is evaluated using metrics aligned with novelty identification and expert time efficiency.

1.3 Research Gap

1.3.1 Discovery-First Evaluation and Real Expert-Time Measurement

Most bioacoustic machine learning studies evaluate performance using standard detection or classification metrics such as F1 score, AUC, or mean average precision. While these metrics are appropriate for recognizing known classes, they do not capture the effectiveness of discovery. In particular, there is no widely adopted evaluation protocol that measures how quickly a system helps experts identify new patterns, how many biologically meaningful discoveries are validated per hour of expert effort, or how much time is spent reviewing irrelevant results. Although some tool-oriented studies report practical speed improvements [2], methodological papers more often focus on reducing label counts rather than measuring actual expert time. This creates a clear gap: discovery-focused metrics and time-based evaluation protocols are missing as standard practice.

1.3.2 No Integrated End-to-End Discovery Pipeline

The literature provides strong components, but they are usually disconnected. Clustering-based discovery can reveal structure, but often lacks active learning and boundary refinement [14, 16]. Active learning can reduce labeling effort, but it is rarely designed to maximize discovery yield of new call types [1, 8, 9]. Weak-to-strong refinement improves temporal boundaries, but it is not typically tied to unsupervised discovery as a mechanism for creating and validating new categories [10]. Tools support annotation and reproducibility, but they do not embed a unified discovery loop with standardized discovery metrics and time logging [2, 3, 20]. As a result, there is no well-established framework that links continuous audio processing, candidate selection, unsupervised discovery, expert-guided querying, boundary refinement, and validation of new categories within a single unified workflow.

1.3.3 Guillemot Long-Term Monitoring Conditions are Under-Validated

Many existing approaches are validated on curated bird datasets, different species, or alternative soundscapes such as coral reefs, whereas guillemot colony recordings present a distinct set of challenges. These recordings are characterized by dense overlapping vocalizations, strong wind and environmental noise, and significant seasonal variation in acoustic patterns [7, 17, 19]. In addition, long-term monitoring requires robustness to acoustic drift, as background conditions and vocal behaviour evolve over weeks and months. AukLab-Audio recordings further introduce a multi-channel setting, yet most current discovery pipelines rely on mono or downmixed audio and do not fully exploit cross-channel information for noise rejection and event verification. Overall, the domain gap is not only species-specific; it is also about long-term drift and multi-channel realism under colony conditions.

1.4 Problem Statement

In long-term monitoring, manual inspection and annotation by domain experts is time-consuming and costly, making it impractical to scale analysis to thousands of hours of audio. Current approaches typically prioritize classification accuracy over discovery yield per expert hour, and methods validated on curated datasets often fail under real guillemot colony conditions, characterized by multi-channel audio (eight high-quality DPA microphones), offering opportunities for cross-channel validation of acoustic events. Most bioacoustic anomaly detection methods are developed for mono recordings and do not exploit spatial redundancy. In dense colony environments, this limitation increases the risk of surfacing channel-specific artifacts rather than biologically meaningful events. Despite advances in bioacoustic machine learning, no integrated framework exists that effectively combines unsupervised anomaly detection for candidate discovery with active learning for expert-efficient validation, specifically optimized for identifying novel vocalizations in continuous field recordings.

This thesis therefore addresses the problem of designing, implementing, and evaluating a discovery pipeline that: (1) identifies rare acoustic candidates via unsupervised methods, (2) optimizes expert annotation effort through active learning, and (3) validates methods for the semi-automatic expansion of acoustic vocabularies by incorporating expert feedback within anomaly detection and active learning loops. Finally, the work aims to create practical tools and workflows that assist human annotators by prioritizing promising audio segments flagged by these intelligent systems, for example by extending an existing audio annotation workflow with the developed discovery and prioritization methods.

1.5 Ethical, societal and sustainability aspects

This research contributes to environmental sustainability by supporting large-scale monitoring of common guillemots, which act as indicators for Baltic Sea ecosystem health. Changes in their vocal behaviour can reflect stress caused by climate change, pollution, overfishing, and habitat disturbance. By enabling more efficient analysis of long-term acoustic data, this work supports earlier detection of ecological change and more informed conservation decisions.

The proposed framework is designed to support, not replace, ecological experts. Machine learning is used only to prioritise data for review, while all interpretation and validation remain human-driven. This reduces repetitive manual effort without removing scientific responsibility. If expert annotation sessions are conducted, interactions will be limited to professional task feedback and time measurements, with all reported results aggregated and anonymized.

All data are collected with institutional approval at the Stora Karlsö research station and contain only environmental sounds and animal vocalisations. No human subjects or personal data are involved. The work will follow applicable BTH/RISE research guidelines, and any audio data remain the property of the data owner (RISE/AukLab) and are used under access restrictions; only non-sensitive derived artifacts will be published.

All methods and tools will be released as open-source software through the AukLab-Audio Initiative, ensuring transparency, reproducibility, and access for researchers and conservation groups worldwide. Although the developed techniques could be misused in surveillance contexts, this thesis explicitly frames them for biodiversity research and commits to open, ethical, and transparent application.

2 Aim and objectives

2.1 Aim

The aim of this thesis is to design, implement, and evaluate a data-centric artificial intelligence framework for acoustic discovery that combines anomaly detection and active learning to help identify rare and novel guillemot vocalisations in long-term recordings. The framework seeks to improve discovery-driven bioacoustic analysis by making the exploration of complex soundscapes more efficient and practical.

2.2 Objectives

- Develop and evaluate anomaly detection methods to identify rare acoustic patterns in long-term guillemot recordings using unsupervised techniques.
- Design and implement active learning strategies that intelligently select informative audio segments to guide efficient annotation and expand knowledge of guillemot vocalizations.
- Integrate anomaly detection and active learning into an iterative discovery pipeline that supports semi-automatic expansion of acoustic categories.
- Experimentally evaluate the proposed framework on real world guillemot recordings to assess its effectiveness in improving discovery and annotation efficiency.

3 Research Questions

RQ1: How effectively can unsupervised, data-driven methods uncover rare or unusual guillemot vocalizations in long, continuous bioacoustic recordings?
Sub-RQ1a: Which sound representations best support this discovery?
Sub-RQ1b: Which structure revealing approaches best organize and separate the acoustic data into interpretable patterns for discovery?

Justification: The AukLab recordings are large, continuous, and mostly unlabeled, so discovery must begin without predefined call categories. A known gap in prior work is that generic anomaly detection often highlights irrelevant noise rather than meaningful biological novelty. This research question tests whether unsupervised discovery is feasible under real guillemot colony conditions and identifies which design choices make it practical.

Sub-RQ1a compares alternative sound representations (e.g., spectrogram-based features and pretrained embeddings) to determine which best preserves relevant acoustic differences in noisy, overlapping colony soundscapes.

Sub-RQ1b evaluates alternative structure-revealing approaches (e.g., clustering and embedding-based exploration versus novelty/anomaly scoring) to determine which methods most effectively separate the soundscape into consistent patterns that can support expert interpretation and further analysis. Together, these results identify the most reliable unsupervised components to carry forward into later stages of the thesis.

RQ2: Which active learning frameworks are most suitable for guiding annotation in long, continuous guillemot recordings when combined with unsupervised discovery methods?

Justification: Expert annotation is time-consuming and costly in long-term monitoring, but active learning in bioacoustics is often evaluated in settings that assume predefined classes and well-structured training data. A key gap is the lack of clear guidance on which active learning frameworks work best when the data is continuous, highly imbalanced, and discovery-oriented (i.e., the goal is to find previously unseen events rather than only improve known-class accuracy). This question focuses on designing and evaluating active learning frameworks that can operate effectively on candidate outputs from unsupervised discovery, and that can prioritize informative segments for expert review in open-ended settings.

RQ3: To what extent does an integrated workflow combining unsupervised discovery and active learning improve discovery efficiency and reduce annotation effort in real guillemot colony recordings?

Justification: The literature contains strong individual components (unsupervised discovery methods, active learning strategies, and annotation-support techniques), but a major gap is the limited evidence on how well these components work when integrated into a single workflow for continuous field recordings. This question evaluates the overall value of integration by assessing whether the combined workflow leads to more efficient discovery and more usable annotations in practice. It provides the end-to end validation needed to argue that a discovery-first, data-centric approach is not only technically feasible but also practically beneficial for long-term guillemot monitoring.

4 Methodology

4.1 Research design

This thesis follows a **design science research** approach. The objective is to design, implement, and evaluate a practical **discovery-based bioacoustic framework** that supports efficient identification of rare or unusual guillemot vocalizations in long, continuous recordings. A design science approach is appropriate because the main contribution of the thesis is a functional artifact in the form of a method and workflow developed to address a real-world problem in long-term ecological monitoring.

The work will be carried out in an iterative manner. In each iteration, candidate methods will be implemented, tested on real AukLab recordings, and refined based on measured outcomes related to discovery performance and annotation efficiency. Feedback from the RISE/AukLab context will be incorporated throughout the process to ensure that the resulting workflow remains practical and relevant for real monitoring applications.

4.2 Data and preprocessing

The primary dataset is provided by RISE through the Stora Karlsö Auk Lab and consists of continuous, multi-channel recordings of common guillemot colonies captured using high-quality microphones. The full archive contains approximately 12 TB of audio data. For practical experimentation, a representative subset of this archive will be used, selected to reflect the diversity and complexity of the full dataset.

The recordings are stored as long continuous audio files with an average duration of approximately 9–10 minutes per file. The dataset reflects realistic field conditions, including dense vocal activity, overlapping calls, environmental noise, and long-term seasonal variation. This scale and complexity make the dataset well suited for evaluating anomaly detection and active learning methods under practical monitoring conditions.

Preprocessing will follow a consistent and reproducible procedure:

- **Segmentation:** continuous audio will be divided into short, overlapping time windows suitable for scalable processing.
- **Spectrogram generation:** time–frequency representations (e.g., log-mel spectrograms) will be extracted using fixed parameter settings.
- **Normalization:** basic normalization will be applied to reduce sensitivity to recording-level variation and improve comparability across time periods.

4.3 Methods for answering the research questions

4.3.1 Unsupervised discovery (RQ1)

To answer RQ1, the thesis will compare unsupervised, data-driven methods for surfacing rare or unusual sounds from long recordings. Two design choices are central:

- **Representations (Sub-RQ1a):** compare discovery performance using different sound representations, including spectrogram-based features and pretrained embeddings.
- **Structure discovery (Sub-RQ1b):** compare approaches that reveal acoustic structure, such as clustering/embedding exploration and novelty/anomaly scoring.

The output of this stage is a ranked and/or grouped set of candidate segments that are likely to contain biologically meaningful rare or unusual vocalizations. It is important to distinguish the scope of this thesis from existing work such as Adaptive Change-Point Detection , which focuses on efficient segmentation and boundary refinement of known vocalization events. In contrast, the present work emphasizes open-ended anomaly detection for surfacing rare or previously uncharacterized acoustic phenomena. The proposed framework complements detection-oriented approaches by prioritizing discovery in largely unlabeled and structurally complex recordings.

4.3.2 Active learning for annotation efficiency (RQ2)

To answer RQ2, active learning strategies will be implemented to prioritize which candidate segments should be reviewed first. The comparison will include commonly used strategy families (e.g., uncertainty-based and diversity-based). The focus is not only on model improvement, but on reducing annotation effort while maintaining discovery output.

4.3.3 Integrated workflow and label refinement (RQ3)

To answer RQ3, the strongest components from RQ1 and RQ2 will be combined into an iterative workflow. The workflow will support repeated cycles of: (i) unsupervised candidate discovery, (ii) prioritized expert review via active learning, and (iii) refinement of labels to improve their usefulness for downstream analysis. The aim is to test whether integration leads to measurable gains compared to conventional practices.

4.4 Evaluation strategy and measurements

The evaluation strategy is designed to reflect the discovery oriented goals of the thesis while maintaining comparability with established machine learning evaluation practices. The framework will be assessed along three complementary dimensions: discovery effectiveness, annotation efficiency, and conventional performance metrics.

Discovery effectiveness. The ability of the system to surface meaningful rare or novel events will be measured using ranking- and discovery-oriented metrics. These include measures such as *Discovery@K* (the proportion of top-ranked candidates validated as relevant events) and novelty yield within selected clusters. Expert validation will be used to confirm whether surfaced candidates correspond to biologically meaningful vocalizations.

Annotation efficiency. Efficiency will be evaluated using time-based measurements that quantify expert effort. Metrics will include discovery yield per hour of review, time-to-first-discovery, and the proportion of reviewed segments judged irrelevant (false discovery burden). These measurements directly capture the practical usefulness of the framework in reducing unnecessary expert workload.

Conventional performance metrics. Where labeled segments are available, standard supervised metrics such as *precision*, *recall*, and *F1-score* will be computed to evaluate detection quality. These metrics provide an additional reference point for comparing the proposed methods with existing approaches and ensure alignment with common evaluation standards in bioacoustic machine learning.

Baseline comparison. The integrated workflow will be compared against practical baselines, including manual expert review, random sampling for annotation, and single-method approaches that use only anomaly detection or only active learning. Comparative evaluation will examine both discovery outcomes and annotation efficiency to determine whether the integrated pipeline provides measurable improvements under realistic monitoring conditions.

5 Expected outcomes

This thesis is expected to deliver a data-centric framework for acoustic discovery in long and continuous guillemot colony recordings. The main outcome will be an experimental workflow that combines unsupervised discovery methods with active learning to help identify rare vocalizations while reducing unnecessary expert effort.

Scientific and technical outcomes. The work is expected to provide (i) a comparative assessment of unsupervised discovery approaches under realistic guillemot colony conditions, including how different sound representations and structure-revealing methods affect discovery performance (RQ1); (ii) evidence on which active learning strategies most effectively reduce annotation effort while maintaining discovery output (RQ2); and (iii) an end-to-end evaluation demonstrating whether an integrated discovery workflow provides measurable gains compared to conventional practices (RQ3). In addition, the thesis is expected to propose discovery-first evaluation measures, such as discovery yield per hour and time-to-first-discovery, which directly address a current gap in bioacoustic machine learning evaluation.

Practical outcomes for stakeholders. For ecologists and the AukLab–RISE monitoring program, the framework is expected to act as a scalable support tool that prioritizes candidate segments likely to contain biologically meaningful rare events, reducing time spent on irrelevant noise and repetitive samples. This can enable more efficient exploration of large sound archives and support earlier identification of changes in colony behavior. For the machine learning community, the work contributes a realistic case study and evaluation protocol for discovery-oriented learning on long, unlabeled field recordings, highlighting what methods work, what fails, and why.

Reproducibility and reuse. The thesis is expected to produce a well-documented, reproducible implementation of the proposed workflow, including scripts and configuration settings that enable replication of experiments. Where permissible, code and non-sensitive artifacts will be released through RISE to support reuse and extension by other researchers working on discovery-driven bioacoustics.

6 Time and activity plan

6.1 Supervision plan

Throughout this thesis project, we will work under the guidance of our university supervisor and in collaboration with supervisors at RISE. Regular weekly supervision meetings will be held, either in person and virtually, to review progress, discuss technical challenges related to anomaly detection and active learning, and receive feedback on implementation and written drafts. In addition to these meetings, we will participate in bi-weekly joint meetings with other students and supervisors working on related guillemot soundscape projects. These group sessions will support knowledge sharing, cross-project discussion, and collaborative problem solving. Continuous communication via email and agreed digital channels will ensure timely follow-up between meetings and help maintain steady progress throughout the thesis.

Table 1: Time and Activity Plan

S.No.	Task	Start	End	Duration
1	Project Proposal Plan Preparation	Jan 20th	Feb 8th	3 weeks
2	Systematic Literature Review (bioacoustics, anomaly detection, active learning)	Jan 20th	Feb 16th	4 weeks
3	Dataset Access & Preprocessing (guillemot recordings, segmentation, feature extraction)	26th Jan	Feb 16th	3 weeks
4	Baseline Anomaly Detection Model Development (autoencoders, one-class methods)	Feb 17th	Mar 9th	3 weeks
5	Evaluation of Anomaly Detection Outputs (noise filtering, rare event relevance analysis)	Mar 10th	Mar 16th	1 week
6	Active Learning Framework Design (query strategies, uncertainty + diversity sampling)	Mar 17th	Mar 30th	2 weeks
7	Integration of Anomaly Detection + Active Learning Loop	Mar 31st	Apr 13th	2 weeks
8	Iterative Acoustic Vocabulary Expansion Experiments (novel call discovery)	Apr 14th	Apr 27th	2 weeks
9	Annotation-Support Tool/Workflow Development (prioritized review interface, ranking pipeline)	Apr 28th	May 5th	1 week
10	Thesis Documentation (ongoing writing, methodology + results updates)	Jan 20th	May 10th	16 weeks
11	**Thesis Draft Submission**	**May 10th**	**Deadline**	**Submission**
12	Opposition Report Preparation & Submission	May 11th	May 24th	2 weeks
13	Presentation Preparation & Thesis Defense	May 18th	May 29th	2 weeks
14	**Final Documentation & Thesis Submission**	**Jun 7th**	**Deadline**	**Submission**

7 Limitations and Risk Management

This section outlines the key limitations of the proposed thesis and examines potential risks that could influence the successful completion of the project. For each identified risk, mitigation strategies are proposed to reduce its impact.

7.1 Limitations

- **Data Annotation Scarcity:** This thesis focuses on discovery within largely unlabeled long-term recordings, which reflects realistic ecological monitoring conditions. However, the limited availability of expert annotations may restrict the scope of quantitative evaluation and benchmarking. To address this, the study combines expert validation, qualitative assessment, and discovery-oriented evaluation metrics instead of relying solely on large fully labeled datasets.
- **Ground Truth Ambiguity for Novel Events:** Rare or previously undocumented vocalizations may not have clearly defined ground-truth categories. This can introduce a degree of subjectivity in validation and may affect reproducibility. The thesis addresses this limitation by applying structured expert review procedures and maintaining transparent documentation of all discovered acoustic patterns.
- **Computational Constraints:** Processing long multi-channel recordings and training embedding-based anomaly detection models requires significant computational resources. These constraints may limit the scale of experiments or the number of model configurations that can be explored within the available timeframe.
- **Generalization Across Environments:** The proposed framework is developed and evaluated primarily on guillemot colony recordings. Although the methods are intended to be transferable, applying them to other ecological environments may require adaptation and additional validation.
- **Evaluation Scope:** Because the thesis emphasizes discovery efficiency and novelty detection rather than traditional classification accuracy, direct comparisons with standard supervised benchmarks may be limited. The evaluation therefore focuses on metrics that better reflect real-world discovery workflows.

7.2 Risk Management

Table 2 summarizes the major risks associated with the thesis and the corresponding mitigation strategies.

Table 2: Risk Management Plan

Risk	Impact	Mitigation Strategy
Anomalies are dominated by artifacts (wind, Sea Waves)	High	Add filtering steps, check consistency across multiple audio channels, and use expert feedback to improve how candidate sounds are selected
Limited availability of expert validation time	High	Use active learning strategies that reduce expert workload by selecting short audio clips and organizing structured annotation sessions to improve efficiency.
High computational cost of large-scale audio processing	High	Use pretrained embeddings and efficient batch processing, and rely on shared GPU resources to support incremental and efficient experimentation.
Uncertainty in measuring discovery performance	Medium	Define clear time-based and discovery-focused metrics, combine quantitative results with expert qualitative assessment, and clearly document the evaluation process.
Integration complexity between system components	Low–Medium	Develop modular pipeline components, test each module independently, and maintain reproducible experiment configurations.

References

[1] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza, "Power to the people: The role of humans in interactive machine learning," *AI Magazine*, 2014. [Online]. Available: https://onlinelibrary.wiley.com/doi/epdf/10.1609/aimag.v35i4.2513

[2] L. authors, "Leaves: an interactive tool that accelerates ecoacoustic data annotation," *Ecological Informatics*, 2025, 10.1016/j.ecoinf.2025.103026. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1574954125000354

[3] W. authors, "Whombat: An open-source audio annotation tool for machine-learning-assisted bioacoustics," *Methods in Ecology and Evolution*, 2025, 10.1111/2041-210X.14468. [Online]. Available: https://arxiv.org/abs/2308.12688

[4] B. Bayram, M. H. Sazlı, *et al.*, "An incremental class-learning approach with acoustic novelty detection for acoustic event recognition," *Sensors*, 2021, 10.3390/s21196622. [Online]. Available: https://www.researchgate.net/publication/355222558_An_Incremental_Class-Learning_Approach_with_Acoustic_Novelty_Detection_for_Acoustic_Event_Recognition

[5] Y. Chen, *et al.*, "Acoustic detection of unknown bird species and individuals," *CAAI Transactions on Intelligence Technology*, 2021, 10.1049/cit2.12007. [Online]. Available: https://www.researchgate.net/publication/349991433_Acoustic_detection_of_unknown_bird_species_and_individuals

[6] V. Dumoulin, O. Stretcu, J. Hamer, L. Harrell, R. Laber, H. Larochelle, B. van Merriënboer, A. Navine, P. Hart, B. Williams, T. A. C. Lamont, T. B. Razak, M. C. R. Team, S. Brodie, B. Doohan, P. Eichinski, P. Roe, L. Schwarzkopf, and T. Denton, "The search for squawk: Agile modeling in bioacoustics," *arXiv*, 2025,

10.48550/arXiv.2505.03071. [Online]. Available: https://arxiv.org/abs/2505.03071

[7] J. Hentati-Sundberg, *et al.*, "Technological evolution generates new answers and new ways forward: A progress report from the first decade at the karlsö auk lab," *Marine Ornithology*, 2025, 10.5038/2074-1235.53.1.1612. [Online]. Available: https://chooser.crossref.org/?doi=10.5038%2F2074-1235.53.1.1612

[8] A. Kholghi, *et al.*, "Active learning for classifying long-duration audio recordings of the environment," *Methods in Ecology and Evolution*, 2018, 10.1111/2041-210X.13042. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/2041-210X.13042?utm_medium=article&utm_source=researchgate.net

[9] J. Lindholm, *et al.*, "Aggregation strategies for efficient annotation of bioacoustic sound events using active learning," *EUSIPCO / arXiv*, 2025, arXiv:2503.02422. [Online]. Available: https://arxiv.org/abs/2503.02422

[10] J. Martinsson and O. Mogren, "Adaptive change-point detection with active learning for sound event labels," *EUSIPCO*, 2024, (see EUSIPCO 2024 paper). [Online]. Available: https://arxiv.org/abs/2403.08525

[11] ——, "Improving active learning efficiency via hybrid disagreement-diversity sampling strategies," *EUSIPCO*, 2025, eUSIPCO 2025.

[12] B. J. McEwen, *et al.*, "An improved computational bioacoustic monitoring approach for detecting sparse features," *Proc. Meetings Acoustics / ASA*, 2024, 10.1121/2.0001837. [Online]. Available: https://www.researchgate.net/publication/376016310_An_improved_computational_bioacoustic_monitoring_approach_for_detecting_sparse_features

[13] B. B. McEwen, "Computational bioacoustics for the detection of rare acoustic events," *PhD Thesis*, 2024, (thesis). [Online]. Available: https://ir.canterbury.ac.nz/items/e27777ce-148b-443d-819d-8f09974510c0

[14] E. Ozanich, A. Thode, P. Gerstoft, L. A. Freeman, and S. Freeman, "Deep embedded clustering of coral reef bioacoustics," *JASA*, 2021, 10.1121/10.0004221. [Online]. Available: https://arxiv.org/abs/2012.09982

[15] E. Ozanich, A. Thode, P. Gerstoft, and S. Freeman, "Unsupervised clustering of coral reef bioacoustics," *arXiv*, 2020, arXiv:2012.09982. [Online]. Available: https://arxiv.org/abs/2012.09982

[16] Y. F. Phillips, M. Towsey, and P. Roe, "Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation," *PLOS ONE*, 2018, 10.1371/journal.pone.0193345. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0193345

[17] RISE, "Rise: Vi lyssnar på sillgrisslor med ai för att bättre förstå biologisk mångfald (project news)," *RISE news*, 2025.

[18] B. Settles, "Active learning literature survey," *University of Wisconsin, Computer Sciences Technical Report*, 2009. [Online]. Available: https://minds.wisconsin.edu/handle/1793/60660

[19] A.-A. team, "Auklab audio research (project description and system design)," *Project page*, 2025. [Online]. Available: https://auklab-audio.github.io/

[20] O. team, "A collaborative open-source tool for reproducible bioacoustic analysis," *Methods in Ecology and Evolution*, 2023, 10.1111/2041-210X.14196. [Online].

Available: https://www.researchgate.net/publication/373027647_OpenSoundscape_An_open-source_bioacoustics_analysis_package_for_Python

[21] M. Towsey, *et al.*, "Long-duration, false-colour spectrograms for detecting species in large audio data-sets," *Journal of Ecoacoustics*, 2018, 10.22261/JEA.IUSWUI. [Online]. Available: https://jea.jams.pub/article/2/1/48

[22] B. van Merriënboer, V. Dumoulin, J. Hamer, L. Harrell, A. Burns, and T. Denton, "Perch 2.0: The bittern lesson for bioacoustics," *arXiv / Google Research*, 2026, arXiv:2508.04665. [Online]. Available: https://arxiv.org/abs/2508.04665

[23] C. M. Wood, *et al.*, "The machine learning–powered birdnet app reduces data deficiency in global bird monitoring," *PLOS Biology*, 2022, 10.1371/journal.pbio.3001670. [Online]. Available: https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001670

A Signature enclosure

The company/ organization *RISE Research Institutes of Sweden* accepts the description of the “work” in this document (“Project plan”). *RISE Research Institutes of Sweden* also accepts to offer the student/ students a reasonable amount of supervision in connection with the future degree project in relation to the parts of the degree project that are related to the ”work” at *RISE Research Institutes of Sweden*. RISE Research Institutes of Sweden is aware that BTH is not part of any agreement with the company and does not guarantee that the “work” is carried out in a satisfactory manner.

The part of the degree project that is carried out by the student/ students “on behalf of” *RISE Research Institutes of Sweden* is only a commitment between the student/ students and the com pany. BTH will evaluate the work in the form of documents that are submitted for evaluation. All submitted works can be made public on request. According to regulations, the passed final report will be published in the document management system “DIVA” and as a part of this publication will then become a generally accessible public document.

This signature enclosure is not a legal agreement and is therefore not binding for any part (BTH nor *RISE Research Institutes of Sweden*) but should be considered as a “declaration of intention” *RISE Research Institutes of Sweden*

Karlskrona, February 8, 2026

____

Signature by supervisor at RISE Research Institutes of Sweden

____

Printed name