Assessing a Domain-Adaptive Deployment Workflow for Selective Audio Recording in Wildlife Acoustic Monitoring

Julia Azziz[1], *Josefina Lema*[1], *Maximiliano Anzibar Fialho*[1], *Lucía Ziegler*[1], *Leonardo Steinfeld*[1], *Martín Rocamora*[1,2]

[1]Universidad de la República, Montevideo, Uruguay [2]Universitat Pompeu Fabra, Barcelona, Spain

***Abstract*—Passive acoustic monitoring is a valuable tool for wildlife research, but scheduled recording often results in large volumes of audio, much of which may not be of interest. Selective audio recording, where audio is only saved when relevant activity is detected, offers an effective alternative. In this work, we leverage a low-cost embedded system that implements selective recording using an on-device classification model and evaluate its deployment for detecting penguin vocalization. To address the domain shift between training and deployment conditions (e.g. environment, recording device), we propose a lightweight domain adaptation strategy based on fine-tuning the model with a small amount of location-specific data. We replicate realistic deployment scenarios using data from two geographically distinct locations, Antarctica and Falkland Islands, and assess the impact of fine-tuning on classification and selective recording performance. Our results show that fine-tuning with location-specific data substantially improves generalization ability and reduces both false positives and false negatives in selective recording. These findings highlight the value of integrating model fine-tuning into field monitoring workflows, in order to improve the reliability of acoustic data collection.**

***Index Terms*—domain shift, bioacoustics, passive acoustic monitoring**

1. INTRODUCTION

Passive acoustic monitoring (PAM) has become an essential tool in ecological and environmental research. The deployment of autonomous recording units (ARUs) enables long-term, non-invasive monitoring of wildlife and ecosystems. Technological advances have made remote sensing and monitoring tools increasingly accessible and efficient [1], [2], reducing fieldwork time, minimizing ecosystem disturbance, and significantly lowering operational costs [3]. These tools also allow data collection at broad spatial and temporal scales, facilitating more detailed and extensive studies of ecological change. This is particularly valuable in regions where extreme conditions and remoteness pose challenges for direct observation, as is the case for penguin colonies.

Penguins have been widely recognized as key bioindicators for monitoring the health of marine ecosystems in the Southern Hemisphere [4], [5]. In the Antarctic region, species such as the Adélie and Gentoo penguins play a vital ecological role. These species breed in colonies distributed along the Antarctic Peninsula and nearby islands, such as Ardley Island, an Important Bird Area and a protected site under the Antarctic Treaty System. As mesopredators, penguins reflect changes in prey availability, primarily krill, and allow us to infer alterations in the structure and functioning of the Antarctic marine ecosystem [6], [7]. However, significant knowledge gaps remain regarding the plasticity of their annual cycle. Some phases of the cycle have been poorly documented due to the logistical challenges of monitoring, especially in remote and hard-to-access environments. Since phenology varies considerably between colonies and across years, it's essential to implement systematic long-term monitoring that can accurately capture annual variation in the timing of all reproductive cycle phases.

Traditional PAM systems operate on fixed schedules, recording at regular intervals regardless of acoustic activity [8], [9]. While this approach maximizes data coverage, it often results in large volumes of audio, much of which may be irrelevant or silent. This creates challenges in terms of storage, power consumption, and manual analysis. To address these limitations, we designed a selective recording scheme using an audio recorder developed by our team, introduced in [10]. This device runs an embedded classification model, based on a convolutional neural network, and records data only when relevant sound is detected.

Recent advances in deep learning have significantly improved the ability to automatically detect and classify bioacoustic signals [11]. In particular, convolutional neural networks trained on spectrogram representations have become a common and effective approach for bird and animal sound classification [11], and can now be deployed on resource-constrained hardware. This enables real-time inference and automatic on-site sound event detection, demonstrated by lightweight implementations such as *TinyChirp* [12]. In turn, this allows for the development of smart audio recording tools. For example, frameworks such as *acoupi* [13] provide open-source infrastructure for deploying bioacoustic-related machine learning models on embedded devices.

However, while selective recording offers substantial advantages, model performance is often limited by domain shift [14], [15]. Differences in background noise, species behavior or recording hardware between training data and the conditions at the deployment site can hinder network performance [16]. Recent work in domain adaptation has explored strategies to mitigate these challenges, including transfer learning [17], adversarial adaptation [18] and self-training methods [19]. We propose addressing domain shift through lightweight domain adaptation, fine-tuning the base model using a small amount of local data. This approach has been shown to improve performance in bird sound classification when compared to models trained solely on non-local data [11], [20]–[22]. By incorporating even a small amount of location-specific audio samples, fine-tuning helps the model adjust better to local acoustic conditions.

This work is motivated by the need for realistic deployment strategies in remote monitoring campaigns, involving ARUs left in the field for extended periods of time. We focus on the scenario where a pre-trained model is deployed to a brand new location and later fine-tuned using a small amount of location-specific data, collected automatically by the same device over a very brief interval. This reflects a practical workflow in which researchers install recorders, retrieve preliminary audio samples, fine-tune the model offline and then re-deploy the updated system for improved selective recording performance during the remainder of the monitoring campaign.

The key contributions of this work are three: (i) we present a low-power selective recording system capable of real-time sound classification tailored for remote acoustic monitoring; (ii) we evaluate a lightweight domain adaptation strategy to improve performance in new deployment conditions using minimal local data; and (iii) we demonstrate a practical workflow for fine-tuning and redeploying the system in the field. To enable reproducibility and further research, the training and fine-tuning pipeline is openly available[1], and all annotated datasets used for training and evaluation are shared via Zenodo[2].

[1] https://github.com/juliaazziz/aurora-domain-adaptation

[2] https://zenodo.org/records/17121978

The remainder of this paper is organized as follows. Section 2 introduces the context and motivation of the application on which this work is based. Section 3 describes the datasets used for training and evaluation, including both public sources and field recordings. Section 4 outlines the followed methodology, describing the employed model architecture, features and domain adaptation strategy. Section 5 presents experimental results, analyzing the impact of fine-tuning, the effect of varying the number of local samples and the performance of the proposed selective recording system. Finally, Section 6 concludes the paper and discusses the implications of our main findings.

2. APPLICATION OVERVIEW

2.1. Acoustic monitoring

Environmental monitoring in Antarctica is essential for detecting and managing the impacts of human activity and climate change in one of the world's last near-pristine ecosystems [23], [24]. Given the continent's remoteness, logistical challenges and strict environmental protection under the Antarctic Treaty System, remote sensing technologies have become increasingly important [1], [2]. Autonomous systems such as satellite platforms, trap cameras, and acoustic recorders enable long-term, low-impact data collection across vast and inaccessible regions. Acoustic monitoring, in particular, offers powerful and non-invasive tools for tracking biodiversity, human activity, and environmental change [25], [26]. It allows researchers to detect and monitor vocal species, as well as anthropogenic noise, providing insights into ecosystem dynamics and disturbance levels. The Protocol on Environmental Protection to the Antarctic Treaty (1991) encourages long-term environmental monitoring and the use of innovative, low-impact methods. Recent initiatives aim to harmonize and expand remote monitoring networks, including acoustic systems, to better support continent-wide environmental management [27], [28].

2.2. Selective recording workflow

Acoustic wildlife monitoring in remote areas faces several challenges. When systems are deployed in isolated or protected areas, they must operate autonomously for extended periods of time with limited access for maintenance. This places strict constraints on energy consumption and data storage. Traditional approaches, such as continuous or scheduled recording, consume substantial energy and generate large volumes of data, most of which are irrelevant when the goal is to monitor specific species with irregular vocal activity. These methods often result in hours of recordings with no target sounds, leading to inefficient power use and significant need for post-processing.

To overcome these limitations, our approach avoids both continuous and duty-cycled recording modes. Instead, it employs an always-listening, low-power acoustic trigger based on a convolutional neural network trained to recognize vocalizations of the target species. This model continuously scans the incoming audio for relevant patterns and activates the recording mechanism only when a match is detected.

Existing PAM devices, such as AudioMoth [29] and other commercially available recorders, lack the computational capacity to run inference on embedded models. These devices are typically designed for scheduled or threshold-based recording and do not support the continuous inference required for selective recording. To enable real-time, on-device classification, we developed a custom device that uses a hardware accelerator [30]. This allows the system to remain energy-efficient while leveraging an embedded neural network, and making it suitable for deployment in remote, battery-powered scenarios.

Figure 1 illustrates a possible deployment workflow that leverages a selective recorder and integrates a domain adaptation stage using local data. Upon initial deployment, the device collects a small amount of audio samples, which are then manually labeled and used to fine-tune the classification model offline. The device is then re-deployed with the updated model weights, and is left for an extended period of time with improved detection performance.

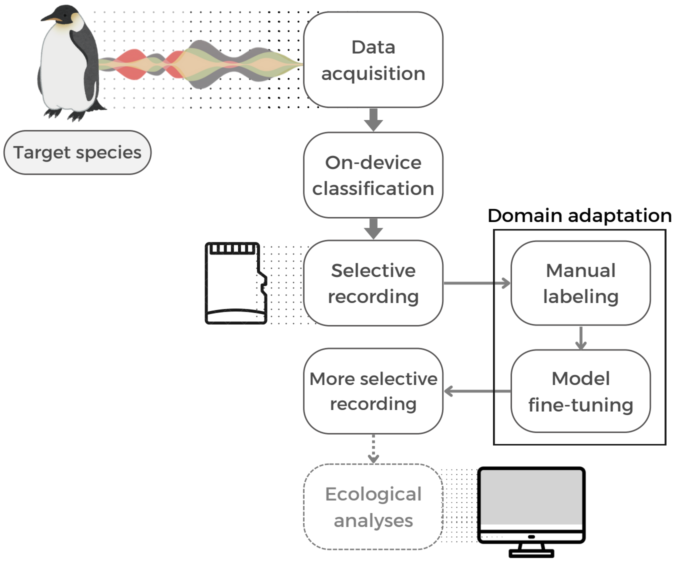

Fig. 1: Diagram of the proposed domain-adaptive deployment workflow for acoustic monitoring.

3. DATA

To train the base model we construct a large training dataset composed of public and controlled sources. We then fine-tune and evaluate the model on a limited amount of controlled audio samples from two geographically distinct Gentoo penguin colonies, which serve as separate locations to test the proposed domain adaptation strategy.

3.1. Training data

The training and validation datasets consist of 1-second audio clips, each one down-sampled to 16 kHz and labeled as either *penguin* or *not penguin*. The positive class examples were sourced primarily from the xeno-canto repository [31], focusing on annotated penguin vocalization recordings, while negative examples were mostly extracted from the FSD50K dataset [32], selecting clips without bird sounds. Additionally, a small portion of the samples were recorded by team members using AudioMoth devices in Ardley, Antarctica, as part of an in-house data collection project. These recordings were manually labeled, resulting in a total of approximately 10 hours for each class, evenly divided, with 80% used for training.

3.2. Target domains

To assess domain generalization and the impact of location-specific fine-tuning, we use three field datasets from two different Gentoo penguin colonies, collected and manually labeled by team members.

Dataset A: this dataset consists of 17 minutes of field recordings from a Gentoo penguin colony in Yorke Bay, Falkland Islands, collected using our device. Of these, approximately 9.1 minutes contain identifiable penguin vocalizations. This dataset represents an unseen domain both in terms of acoustic environment and recording hardware.

Dataset B: this dataset includes 14 minutes of field recordings of Gentoo penguin vocalizations in Ardley Island, Antarctica, of which 6.5 minutes contain penguin vocalizations. For this dataset, samples were recorded using AudioMoths. These samples share many characteristics with a portion of the original training set, as they were recorded in a similar Antarctic environment using the same device.

Dataset C: this dataset contains 12 hours of uninterrupted audio recordings with annotated vocalization intervals, collected using AudioMoths at Yorke Bay. This allows us to replicate continuous audio input in order to evaluate our selective recording system end-to-end.

The first two datasets enable us to contrast domain adaptation performance in two distinct scenarios: one involving a fully novel domain (Dataset A), and another that partially overlaps with the training conditions in both location and recording device (Dataset B).

4. METHODOLOGY

We aim to evaluate whether lightweight fine-tuning on a small amount of location-specific data improves the performance of an embedded classification model deployed in new, unseen environments. To this end, we first train a base classification model on a dataset that excludes recordings from the target domains. We then fine-tune this pre-trained model using a subset of recordings from each target domain and evaluate its performance on held-out data from the same location.

4.1. Audio preprocessing

Each 1-second audio sample is transformed into a log-Mel spectrogram, using a frame size of 30 ms and a hop size of 10 ms for the Short-Time Fourier Transform. We also set the number of Mel bins to 96, resulting in spectrograms of 98 by 96. Features are then quantized to 8-bit integers, simulating embedded inference conditions. Two exemplar spectrograms from Dataset A are shown in Figure 2.

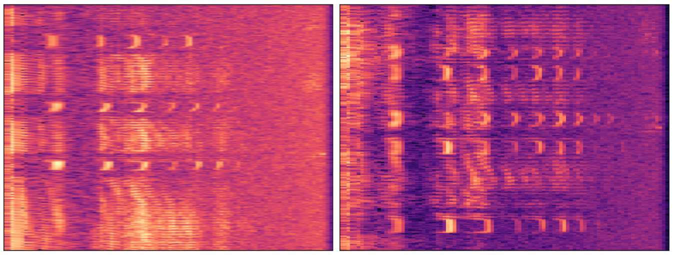

Fig. 2: Two exemplar spectrograms of penguin vocalizations from Dataset A, recorded in the Gentoo colony at Yorke Bay, Falkland Islands.

4.2. CNN architecture

For the classification task, we use a modified version of ResNet18 [33], where the number of filters in each convolutional layer is reduced to one quarter of the original in order to decrease memory and computation requirements. A full description of the modified architecture can be found in Table 1. The final layer outputs a binary classification, indicating the presence of penguin vocalizations. On the recording device, the CNN is fully quantized to 8-bit integers for efficient real-time inference. However, all training, fine-tuning, and evaluation are performed using 32-bit floats on a desktop machine.

The original training is performed for 50 epochs with an initial learning rate of 1×10^{-4}, applying early stopping. We use a scheduler to dynamically reduce the learning rate by a factor of 0.6 when the validation loss plateaus, with a patience of 3 epochs.

4.3. Domain adaptation strategy

To adapt the model to a new environment, we fine-tune the pre-trained base model using a small subset of data from the target domain. Each location-specific dataset is randomly split into 30% for fine-tuning and 70% for validation. Fine-tuning is performed for 15 epochs with an initial learning rate of 1×10^{-4}, and early stopping is applied to prevent overfitting, given the reduced size of the training set. We use a scheduler to dynamically reduce the learning rate by a factor of 0.6 when the validation loss plateaus, with a patience of 3 epochs. We also use 5-fold cross-validation to select the optimal number of layers to unfreeze during fine-tuning. To ensure robustness and mitigate bias introduced by data selection, we repeat this process over 20 splits.

Table 1: Modified ResNet18 architecture used for all experiments. Each row represents the sequence of layers, with **c** channels and stride **s**.

Input	Operator	c	s
$98 \times 96 \times 1$	Conv2D	16	2
$49 \times 48 \times 16$	MaxPool2D	16	2
$25 \times 24 \times 16$	Residual block ×2	16	1
$25 \times 24 \times 16$	Conv2D + Residual block	32	2
$13 \times 12 \times 32$	Conv2D + Residual block	64	2
$7 \times 6 \times 64$	Conv2D + Residual block	64	2
$4 \times 3 \times 64$	GlobalAvgPool	-	-
$64 \times 1 \times 1$	Dense (ReLU)	32	-
$32 \times 1 \times 1$	Dense (sigmoid)	1	-

4.4. Recording hysteresis

In order to prevent frequent toggling between recording states, the recorder device implements a simple hysteresis strategy: audio is recorded when at least $n = 2$ consecutive frames indicate positive detection, and stopped after $m = 2$ consecutive negatives. Part of assessing the effectiveness of domain adaptation is evaluating whether this mechanism also performs better when using the fine-tuned model.

To carry out this analysis, we use Dataset C to replicate continuous audio input, processing the full audio stream using both the baseline and fine-tuned models. The output predictions are used to trigger the hysteresis logic, generating a set of recorded segments. These are then compared against the annotated vocalization intervals to compute false positives and false negatives, as well as precision and recall.

5. EXPERIMENTAL RESULTS

5.1. Impact of domain adaptation

We begin by evaluating how domain-specific fine-tuning impacts the generalization ability of the classification model. For each target dataset, we first assess the performance of the baseline model on the validation split. We then fine-tune this model using the local training data and re-evaluate its performance on the same validation split.

Results for the base and fine-tuned models on both locations are shown in Table 2, averaged across all repetitions. These results show that fine-tuning consistently improves performance across both datasets, yielding gains of up to 20.8% in terms of accuracy and 15.2% in terms of F1-score. Standard deviations across repetitions remain low, indicating that performance gains are stable and robust.

Table 2: Results obtained for the base and fine-tuned models on Datasets A and B, averaged across all repetitions and using 30% of the dataset for training.

	Dataset A		**Dataset B**	
Model	Accuracy (%)	F1-score (%)	Accuracy (%)	F1-score (%)
Baseline	67.8 ± 0.6	74.7 ± 0.8	88.3 ± 0.5	87.9 ± 0.6
Fine-tuned	88.6 ± 0.5	89.9 ± 0.4	95.9 ± 0.4	95.4 ± 0.4

We observe that performance on Dataset B is consistently better than on Dataset A, even before fine-tuning is applied, likely due to two factors. First, Dataset B was recorded using the same device as part of the original training data, reducing variability introduced by differing hardware. Second, although Dataset B was collected independently, it shares environmental context with the training set, which included Antarctic samples from previous field campaigns.

5.2. How many local samples are needed?

The previous section showed that simply fine-tuning on a reduced amount of location-specific samples can improve model performance. However, when local data is particularly limited and must be collected during deployment, the question arises: how many local samples are actually required to noticeably improve model performance?

This analysis is a crucial consideration for assessing the practicality of the proposed deployment workflow, since collecting large volumes of labeled local data may be infeasible due to logistical constraints.

To assess this aspect, we examine the relationship between the number of fine-tuning samples and the improvement achieved compared to the base model. We simulate different data availability scenarios by fine-tuning using varying percentages of the dataset, starting from 5% and up to the 30% employed in the previous section.

Figure 3 shows the results of fine-tuning the model using different amounts of local data from Datasets A and B. All models were evaluated on the full validation set. We observe that even small amounts of local data increase performance and help overcome non-fine-tuned models, particularly in the more challenging conditions of Dataset A. With as little as 30 seconds of training data, the classifier achieves a significant portion (90% or more) of the overall performance gain observed when fine-tuning with 30% of all available data.

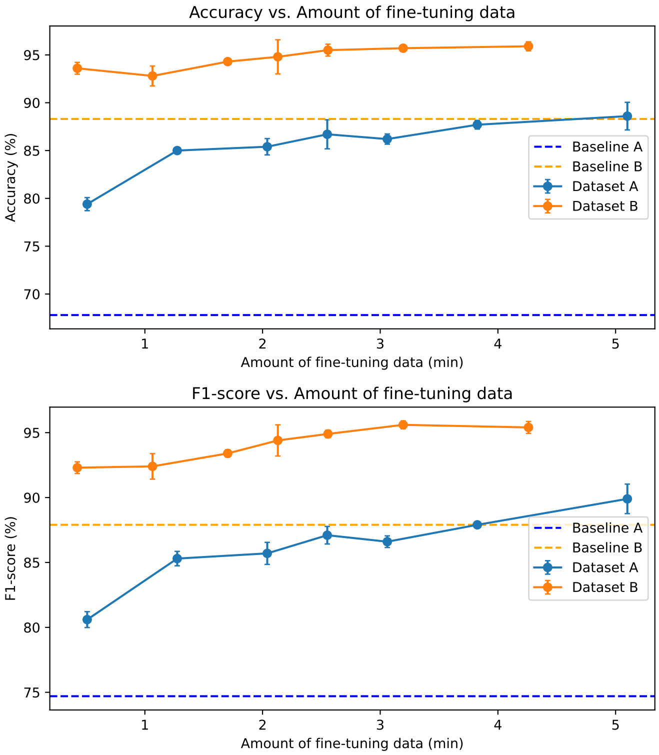

Fig. 3: Results of fine-tuning the model using different amounts of local data from Datasets A and B.

For Dataset B, the accuracy and F1-score curves begin to plateau beyond the 2-minute mark, suggesting diminishing returns as more data is added. This trend indicates that a relatively small amount of labeled local data may be sufficient to adapt a pre-trained model to a location that shares certain characteristics with the original training dataset. For Dataset A, however, performance continues to improve more steadily with increasing amounts of fine-tuning data, particularly in terms of the F1-score, and no clear plateau is observed within the tested range. The contrast between the two datasets highlights the importance of acoustic similarity: when the target environment closely resembles training conditions, minimal fine-tuning may suffice, whereas more distinct domains benefit from additional adaptation data.

5.3. Recording hysteresis

To further evaluate the benefits of domain adaptation, we compare the effectiveness of the recording hysteresis mechanism using the baseline and fine-tuned models. Both models are evaluated on Dataset C, replicating a continuous audio stream. Since the audio samples from said dataset were collected at the Yorke Bay colony, for the fine-tuned model we use the one re-trained on Dataset A.

Table 3 shows the results obtained for both models, presenting the number of false positives, false negatives, true positives and F1-score, all measured in total minutes of audio recorded from the full 12 hours.

Table 3: Results obtained for the base and fine-tuned models on the 12 hours of Dataset C, replicating a recording scheme with the hysteresis mechanism.

Model	True positives	False positives	False negatives	F1-score
Baseline	4.3 min	14.8 min	2.8 min	32.5%
Fine-tuned	5.4 min	7.9 min	1.7 min	52.7%

While both models produce a considerable number of false positives, the true positives rate remains high. In addition, the percentages of true negatives is extremely high, given that, out of 12 hours, less than twenty minutes are recorded. This is particularly important in selective recording, where erroneous recordings would result in higher volumes of irrelevant data. Moreover, the fine-tuned model achieves a better performance, increasing the F1-score value by a net gain of 20.2%. While this is mainly due to the reduction of false negatives, it also responds to a slight rise in true positives. Despite the F1-score value not being particularly high in absolute terms, the relative improvement still highlights the effectiveness of domain adaptation in this context. Overall, these results support the value of incorporating localized data during training to enhance performance in novel field conditions.

6. CONCLUSION

In this work, we presented a domain-adaptive deployment workflow for selective audio recording in wildlife acoustic monitoring. By combining a low-power embedded recording device with an on-device classification model, our system enables energy-efficient monitoring in remote environments such as penguin colonies. To mitigate the challenges introduced by domain shift, we proposed a lightweight fine-tuning strategy based on a small amount of location-specific data.

Through experiments on field data, collected from two geographically distinct Gentoo penguin colonies, we show that domain-specific fine-tuning leads to substantial gains in classification performance. Notably, we find that even small amounts of local data (as little as 30 seconds) can yield meaningful improvements over the base model while requiring minimal annotation effort. This result is particularly relevant for remote monitoring deployments, where collecting large amounts of labeled data during the initial stages is often infeasible.

We also demonstrate that these improvements translate into better selective recording performance, reducing false positives and false negatives in realistic deployment conditions. Future work will focus on validating the proposed domain adaptation strategy in future recordings campaigns in Ardley, Antarctica, as part of an ongoing research and data collection project. This will involve effectively integrating the proposed monitoring strategy into the regular campaign workflow, including the mid-deployment fine-tuning stage with a few labeled local samples.

REFERENCES

[1] W. Turner, S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger, "Remote sensing for biodiversity science and conservation," *Trends in Ecology & Evolution*, vol. 18, no. 6, pp. 306–314, 2003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169534703000703

[2] N. Pettorelli, H. Nagendra, R. Williams, D. Rocchini, and E. Fleishman, "A new platform to support research at the interface of remote sensing, ecology and conservation," *Remote Sensing in Ecology and Conservation*, vol. 1, no. 1, pp. 1–3, 2015. [Online]. Available: https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1002/rse2.1

[3] A. Schulz, C. Shriver, S. Stathatos, B. Seleb, E. Weigel, Y.-H. Chang, B. Saad, D. Hu, and J. Mendelson, "Conservation tools: the next generation of engineering–biology collaborations," *Journal of The Royal Society Interface*, vol. 20, 2023.

[4] P. D. Boersma, "Penguins as marine sentinels," *BioScience*, vol. 58, no. 7, pp. 597–60, 2008.

[5] T. Carpenter-Kling, J. M. Handley, M. Connan, R. J. M. Crawford, A. B. Makhado, B. M. Dyer, W. Froneman, T. Lamont, A. C. Wolfaardt, M. Landman, M. Sigqala, and P. A. Pistorius, "Gentoo penguins as sentinels of climate change at the sub-antarctic prince edward archipelago, southern ocean," *Ecological Indicators*, vol. 101, pp. 163–172, 2019.

[6] F. Cusset, P. Bustamante, A. Carravieri, C. Bertin, and R. Brasso, "Circumpolar assessment of mercury contamination: the adélie penguin as a bioindicator of antarctic marine ecosystems," *Ecotoxicology*, vol. 32, p. 1024–1049, 2023.

[7] A. L. Machado-Gaye, A. Kato, M. Chimienti, N. Gobel, Y. Ropert-Coudert, A. Barbosa, and Álvaro Soutullo, "Using latent behavior analysis to identify key foraging areas for adélie penguins in a declining colony in west antarctic peninsula," *Marine Biology*, vol. 171, 2024.

[8] L. Sugai, C. Desjonquères, T. Silva, and D. Llusia, "A roadmap for survey designs in terrestrial acoustic monitoring," *Remote Sensing in Ecology and Conservation*, vol. 6, 11 2019.

[9] L. S. M. Sugai, C. Desjonquères, T. S. F. Silva, and D. Llusia, "A roadmap for survey designs in terrestrial acoustic monitoring," *Remote Sensing in Ecology and Conservatio*, vol. 6, pp. 220–235, 2020.

[10] J. Azziz, J. Lema, L. Steinfeld, E. Acevedo, and M. Rocamora, "Selective audio recording device for wildlife research using embedded machine learning," in *2025 IEEE Latin Conference on IoT (LCIoT)*, 2025, pp. 65–68.

[11] D. Stowell, "Computational bioacoustics with deep learning: a review and roadmap," *PeerJ*, vol. 10, p. e13152, Mar. 2022. [Online]. Available: http://dx.doi.org/10.7717/peerj.13152

[12] Z. Huang, A. Tousnakhoff, P. Kozyr, R. Rehausen, F. Bießmann, R. Lachlan, C. Adjih, and E. Baccelli, "Tinychirp: Bird song recognition using tinyml models on low-power wireless acoustic sensors," 2024. [Online]. Available: https://arxiv.org/abs/2407.21453

[13] A. Vuilliomenet, S. M. Balvanera, O. M. Aodha, K. E. Jones, and D. Wilson, "acoupi: An open-source python framework for deploying bioacoustic ai models on edge devices," 2025. [Online]. Available: https://arxiv.org/abs/2501.17841

[14] A. L. Bidarouni and J. Abeßer, "Towards domain shift in location-mismatch scenarios for bird activity detection," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1267–1271.

[15] J. Liang, I. Nolasco, B. Ghani, H. Phan, E. Benetos, and D. Stowell, "Mind the Domain Gap: A Systematic Analysis on Bioacoustic Sound Event Detection," in *2024 32nd European Signal Processing Conference (EUSIPCO)*, 2024, pp. 1257–1261.

[16] K. Wilkinghoff, T. Fujimura, K. Imoto, J. L. Roux, Z.-H. Tan, and T. Toda, "Handling domain shifts for anomalous sound detection: A review of dcase-related work," 2025. [Online]. Available: https://arxiv.org/abs/2503.10435

[17] B. Ghani, T. Denton, S. Kahl, and H. Klinck, "Global birdsong embeddings enable superior transfer learning for bioacoustic classification," *Scientific Reports*, vol. 13, no. 1, Dec. 2023. [Online]. Available: http://dx.doi.org/10.1038/s41598-023-49989-z

[18] E. Tzeng, J. Hoffman, K. Saenko, and T. Darrell, "Adversarial Discriminative Domain Adaptation," 2017. [Online]. Available: https://arxiv.org/abs/1702.05464

[19] Y. Zou, Z. Yu, B. V. K. V. Kumar, and J. Wang, "Domain Adaptation for Semantic Segmentation via Class-Balanced Self-Training," 2018. [Online]. Available: https://arxiv.org/abs/1810.07911

[20] P. Lauha, P. Somervuo, P. Lehikoinen, L. Geres, T. Richter, S. Seibold, and O. Ovaskainen, "Domain-specific neural networks improve automated bird sound recognition already with small amount of local data," *Methods in Ecology and Evolution*, vol. 13, no. 12, pp. 2799–2810, 2022. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14003

[21] B. Ghani, V. J. Kalkman, B. Planqué, W.-P. Vellinga, L. Gill, and D. Stowell, "Impact of transfer learning methods and dataset characteristics on generalization in birdsong classification," *Scientific Reports*, vol. 15, no. 1, May 2025. [Online]. Available: http://dx.doi.org/10.1038/s41598-025-00996-2

[22] I. Nolasco, S. Singh, V. Morfi, V. Lostanlen, A. Strandburg-Peshkin, E. Vidaña-Vila, L. Gill, H. Pamuła, H. Whitehead, I. Kiskin, F. H. Jensen, J. Morford, M. G. Emmerson, E. Versace, E. Grout, H. Liu, B. Ghani, and D. Stowell, "Learning to detect an animal sound from five examples," *Ecological Informatics*, vol. 77, p. 102258, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S157495412300287X

[23] D. W. H. Walton and J. Shears, "The need for environmental monitoring in antarctica: Baselines, environmental impact assessments, accidents and footprints," *International Journal of Environmental Analytical Chemistry*, vol. 55, pp. 77–90, 1994.

[24] K. A. Hughes, "How committed are we to monitoring human impacts in antarctica?" *Environmental Research Letters*, vol. 5, 2010.

[25] L. Ziegler and A. Soutullo, "Anthropogenic noise in terrestrial antarctica: a short review of background information, challenges and opportunities," *Polar Research*, vol. 43, Apr. 2024. [Online]. Available: https://doi.org/10.1080/17518369.2024.2330203

[26] R. Gibb, E. Browning, P. Glover-Kapfer, and K. E. Jones, "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring," *Methods in Ecology and Evolution*, vol. 10, no. 2, pp. 169–185, 2019. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.13101

[27] Committee for Environmental Protection, "Final report of the XXV Meeting of the Committee for Environmental Protection, ATCM XLV, Helsinki, Finland," Presented at the Antarctic Treaty Consultative Meeting XLV / CEP XXV, Helsinki, Finland, Jun. 2023.

[28] P. Newman, "Developing environmental monitoring approaches in antarctica," Presented at the Committee for Environmental Protection (CEP), ATCM XLII – CEP XXII, Prague, Czech Republic, 2019.

[29] A. P. Hill, P. Prince, J. L. Snaddon, C. P. Doncaster, and A. Rogers, "AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment," *HardwareX*, vol. 6, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468067219300306

[30] Silicon Labs, *MVP Accelerator*, https://docs.silabs.com/gecko-platform/4.1/machine-learning/tensorflow/mvp-accelerator.

[31] xeno-canto Foundation, "xeno-canto: Sharing bird sounds from around the world." [Online]. Available: https://www.xeno-canto.org

[32] E. Fonseca, X. Favory, J. Pons, F. Font, and X. Serra, "FSD50K: an open dataset of human-labeled sound events," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 829–852, 2022.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015. [Online]. Available: https://arxiv.org/abs/1512.03385