Addressing Data-Scarcity in Automated Bioacoustic Monitoring for Global Bird Conservation

Shane Weisz

November 2024

Contents

1 Abstract

As a scalable solution for tracking bird populations, automated bioacoustic monitoring is an extremely useful tool for conservationists. However, traditional deep learning approaches typically require hundreds to thousands of labeled examples for each species, which is challenging for rare or endangered species, particularly in already data-scarce regions as common in the Global South. Leveraging bioacoustic foundation models trained by self-supervised learning is a promising recent avenue, and is in its infancy in being applied to the bird vocalisation classification task by leading research labs including Google. However, there has yet to be a comprehensive investigation into evaluation and improvement for data-scarce contexts in particular, where such models would be most valuable. This research aims to address this gap by evaluating the baseline performance of existing leading bird call classifiers in data-limited contexts, using a curated benchmark dataset of rare species in Southern Africa

as a case study. To improve performance, we will explore a variety of few-shot learning approaches, incorporate contextual metadata such as GPS location and time-of-season, and leverage multi-modal data like satellite imagery for habitat information. The overarching aim is to develop a generalizable approach to improving bird call classification in the face of data scarcity, to support pressing initiatives in biodiversity monitoring and conservation worldwide.

2 Introduction

The ability to monitor bird populations is crucial for conservationists. More generally, bird populations serve as important indicators of overall ecosystem health and biodiversity, providing a real-time measure of the impact of global climate change on the environment. As a result, there are various projects worldwide for tracking bird populations, from long-running global citizen science initiatives like eBird [Sullivan et al., 2009] to regional initiatives like the Southern African Bird Atlas Project [Lee et al., 2022].

Traditional methods for monitoring bird populations, including manual acoustic surveying, tend to be highly labor-intensive and are often impractical for large-scale deployment. *Automated* acoustic monitoring, enabled by the deep learning revolution over the past decade, thus offers a scalable alternative.

Deep learning approaches for bioacoustic classification, however, typically require hundreds to thousands of labeled examples per species. These training examples are usually gathered through a combination of citizen scientists and expert annotators – citizen scientists help with collect recording data at scale, whilst expert annotators can validate the quality of the data.

However, for species for which there is limited labeled training data, these data quantity requirements for traditional deep learning approaches pose a significant challenge. This is particularly relevant to the Global South, where there are comparatively fewer both citizen scientists and expert annotators compared to the much-more studied Northern Hemisphere. Moreover, it is often the case that the more endangered a species is, the more difficult it is to acquire training data (since they are by definition scarcer)– yet these are of course the species for which an automated monitoring solution would be most useful to conservationists!

As a motivating example for where an automated monitoring solution would be extremely useful, consider the case of Botha's Lark, South Africa's most threatened endemic bird species, restricted to a highly localized region of Southern Africa. Botha's Lark has seen a dramatic decline in numbers in recent years, decreasing by over 90% in the last decade alone, with recent data suggesting as few as 340 individuals may be left [1]. This has resulted in BirdLife South Africa (the country's dedicated bird-conservation organisation) implementing the Botha's Lark Species Action Plan for rapid action to protect the species from extinction [2]. One of the techniques used for monitoring the species thus far

[1] See Botha's Lark Observational Guidelines here.
[2] https://www.birdlife.org.za/bothas-lark/

has been manual acoustic surveying by experts, and so an automated approach would clearly be highly useful in scaling this solution. In practice, though, this is limited by the fact that there are only 2 recordings on xeno-canto [3], and 4 in the Macaulay Library database [4], the two primary public dataset sources for labeled bird sound recordings, making a few-shot learning approach the only feasible approach to make the most of any new data that could be acquired.

This case of Botha's Lark presents just one example of the value that research into data-efficient bird call classification could make if applied in the real-world. This thus motivates this proposal for investigation into making the most of limited available training data to produce useful real-world bird sound classification models that can then assist conservationists in automated bioacoustic monitoring efforts in data-scarce environments.

3 Literature Review

3.1 Recent work in Automated Bird-Call Classification

There are currently three leading global systems for bird call classification: *BirdNet* [Kahl et al., 2021], developed by the Cornell Lab of Ornithology; *Perch*, developed by Google Research as published in *Nature* [Ghani et al., 2023], and *BirdAVES* [Hagiwara, 2024], developed at the Earth Species Project, an NGO working on decoding non-human communication using AI.

The most extensively studied model, BirdNet, supports over 6000 species worldwide. However, the system requires hundreds of labeled examples per species for training the ResNet-based model. This poses a major limitation in the Southern African context, where significantly less training data is available compared to the data-rich regions of the Global North. As a result, over 35% of Southern African species remain unidentifiable by BirdNet, whilst the developers have acknowledged even for those African species supported by the model supports performance is significantly worse than the Northern Hemisphere [Kahl, 2024].[5] Additionally, there are several other limitations of BirdNet, as discussed by Pérez-Granados [2023], including the lack of public evaluation datasets for reproducibility and comparison of results to quantify performances. Despite its limitations, BirdNet has been deployed as a mobile application so can already be a highly useful tool to assist citizen-scientists with identifications and data collection in-field.

To improve downstream performance, transfer learning was explored by researchers at the University of Helsinki, whose work demonstrates that, whilst

[3] `https://xeno-canto.org/species/Spizocorys-fringillaris`

[4] `https://media.ebird.org/catalog?taxonCode=botlar1&mediaType=audio&sort=rating_rank_desc`

[5] The full list of 6000 species BirdNet can classify is listed here. These are not split into region. Thus, to identify the proportion of species unidentifiable by BirdNet, I then separately pulled the list of ≈ 950 Southern African species from Wikipedia, and then cross-referenced the two lists based on scientific name. I then manually verified a sample of the unidentifiable list, by verifying that they have low data samples on both xeno-canto and in the Macaulay Library database.

not unexpected, fine-tuning a base neural network with local data is highly effective in significantly reducing the number of labeled examples required for attaining a desired level of accuracy on a small dataset of target species [Lauha et al., 2022].

Following on from BirdNet and transfer learning, recent work has investigated Self-Supervised Learning (SSL) and foundational models, following the impressive trends in language modelling with Large-Language Models (LLMs).

For example, recent research by Yang et al. [2024] at Oxford explored a few-shot learning approach for bird sound classification, leveraging a base SSL-trained audio foundation model to produce general audio feature embeddings that are combined with traditional spectrogram-based features before classification. Their findings show improved performance on a small target dataset. There remains a gap in understanding whether this approach would scale effectively to high-dimensional classification tasks with thousands of species, as BirdNet does. Additionally, other few-shot learning frameworks, such as meta-learning and prototypical networks, have yet to be explored in this context.

Google Research recently released their own SSL-based bioacoustic classifier, Perch, which has shown state-of-the-art results [Ghani et al., 2023]. Perch is a large-scale foundational bioacoustic model trained on over 10k species of bird audio data, producing high-quality feature embeddings that can be applied for few-shot learning to generalize to new domains. Similar work on Self-Supervised Learning and for Few-Shot Bird Sound Classification was also done by Moummad et al. [2024].

Following the release of Perch, follow-up work by Van Merriënboer et al. [2024] at Google DeepMind discussed a variety of relevant topics to bioacoustic modelling with its unique characterists of large amounts of noise, strong class imbalances, and distribution shifts. Their work highlighted one of the main bottlenecks in real-world application of advancements in bioacoustic modelling – the lack of robust evaluation procedures to measure model capabalities, and the need for high-quality downstream benchmark datasets. This gap is one of the contributions this proposed research would help in filling.

Notably, a similar observation inspired the work to produce HawkNet, a regional avian classifier for Canada [Huus et al., 2024]. The authors trained their own CCN ensemble classifier on a regional-specific dataset, produced by bootstrapping Perch embeddings for pre-processing to help create high-quality focal recordings. The resulting model showed significant improvements in performance over the BirdNet and Perch global models, thus demonstrating the utility of regional-specific work. However training a local model from scratch relied on leveraging the vast quantities of data available for Canada, less feasible for the comparatively data-scarce Global South. Moreover, HawkNet omitted around one third of Canadian species from their dataset due to insufficient data – and of course many of these species are those for which the classifier would presumably be most useful.

3.2 Southern African studies

In the context of Southern Africa specifically, which this research will use as a globally representative case study of a data-constrained region, there have only been a few small investigations into automated bird sound classification. For example, one study utilized passive acoustic monitoring to record audio at multiple wetland locations, clustering spectrograms and then matching them to xeno-canto data [Mosikidi et al., 2023]. Other small-scale studies include a study by Nel et al. [2024] who developed a cost-effective sound-based sensor system capable of deployment in diverse wetland ecosystems, tested at Rondevlei Nature Centre in Cape Town, South Africa, plus a study by Doell et al. [2024] that trained Convolutional Neural Networks from scratch on a small dataset of common South African bird species in Kwa-Zulu Natal. Whilst these serve as important proof-of-concepts, there is a signicant lack of larger-scale research to support automated bird call classification for Southern Africa.

4 Research Questions

From the available literature, several interesting research directions emerge under the theme of improving bird call classification for data-scarce contexts. Namely, we propose the following key research questions:

1. Can we develop a generalizable evaluation framework for benchmarking bird call classification performance in data-limited contexts, applicable across diverse regions?

2. Using select endangered species in Southern Africa as a case study, how do leading bioacoustic foundation-model based bird call classifers (BirdNet, Perch, BirdAVES) perform for data-scarce species, and how does their accuracy vary with the availability of training data? And does migratory behavior influence performance, due to the accessibility to higher-availability Northern Hemisphere data for such species?

3. To what extent does regional fine-tuning improve performance over general-purpose models in the Southern African context?

4. How does the Self-Supervised Learning (SSL) few-shot learning framework, as proposed by Yang et al. [2024], scale to real-world numbers of species (almost 1k species in Southern Africa), and how could it be enhanced using embeddings from bird-specific foundation models rather than general audio models?

5. How do other few-shot learning techniques, such as prototypical networks, compare to SSL-based approaches in terms of accuracy, scalability, and robustness?

6. How does classification performance scale with the number of training examples, and how does this vary between species with very distinctive

calls (e.g. the Red-Chested Cuckoo's characteristic 'Piet-my-vrou') compared to those with a large variety of different calls (such as the Southern Boubou)?

7. How effectively can temporal (e.g. time-of-season) and spatial (e.g. GPS location) metadata be incorporated into the few-shot learning framework, to either improve classification accuracy or reduce data quantity requirements? Can such metadata help disambiguate calls that are acoustically similar but contextually distinct?

8. Can performance be further improved and/or data quantity requirements reduced by integrating satellite imagery of GPS locations, to derive habitat information as supplementary metadata?

5 Methodology

To conduct this research, we would follow a two-step process: firstly creating a generalizable evaluation framework and benchmark dataset, followed by a suite of experiments to compare and evaluate various data-efficient modeling approaches.

1. Dataset creation:

   - Firstly, a benchmark evaluation dataset needs to be created for standardised model comparisons.
   - The first step would be integrating with public citizen-science datasets (Xeno-canto, Macauley Library, BirdLasser, iNaturalist), designing a robust pipeline that can be generalized to a variety of locales or contexts.
   - For our Southern African context, we would then consult with local experts to curate a high-quality and diverse real-world test set for a chosen set of endangered species. Such a list of species would be obtained by consultation with BirdLife South Africa as to those for which this solution would be most valuable.
   - For example, for the endangered Botha's Lark as introduced above, we could consult with University of Limpopo's lark-expert ornithologist Dr Derek Engelbrecht, who conducted brilliant research into vocalizations and song flight of the similar Pink-billed Lark [Engelbrecht, 2021].
   - The lessons learned from such consultations will be documented clearly for helping with future work generalizing to broader ecoacoustic contexts.

2. Modelling experiments

- Establish baseline performances of BirdNet, Perch, and BirdAves on the evaluation test-set (aggregate and species-wise precision, recall, F1).
- Fine-tune the foundation models on the Southern African specific dataset. Evaluate extent of improvement in performance.
- Evaluate how the performance of few-shot learning approaches scales with number of training examples.
- Evaluate how temporal and spatial metadata affects results, either as a post-processing step, or embedded as inputs to the model before the final classification layer.
- Evaluate whether multi-modal satellite data can improve performance, either as end-to-end imagery provided as input alongside the audio spectral data, or via an intermediary habitat classification model.

6 Expected Contributions

1. A robust and adaptable framework for benchmarking bird call classification performance in data-scarce regions. This framework will be designed to generalize to diverse ecoacoustic datasets globally, beyond Southern Africa.

2. A curated evaluation dataset focused on endangered bird species in Southern Africa as a case study, providing a representative example of data-limited contexts. This dataset will serve as a reference for validating the evaluation framework and testing the performance and scalability of data-efficient modeling approaches.

3. A comprehensive evaluation of how various few-shot learning techniques can improve performance in data-scarce contexts over the baseline existing leading bird call classifiers (BirdNet, Perch, BirdAVES).

4. Insights into how contextual (spatial and temporal) metadata and multimodal data (e.g., satellite imagery) can improve performance and/or decrease data quantity requirements.

5. Open-sourced pipeline software, dataset, and models to facilitate future research into bioacoustic monitoring in both Southern Africa and other data-scarce regions worldwide.

7 Conclusion

This research aims to address data scarcity challenges in bird call classification through investigation of a variety of few-shot learning techniques. Using endangered species in Southern Africa as a case study, the project will aim to provide a scalable, data-efficient classification framework that can be widely applied for regional bioacoustic monitoring. The outcomes of this work would thus help support conservationists in monitoring bird populations more effectively in data-scarce regions, and contribute valuable resources for future research and practical conservation efforts.

References

M. Doell, D. Kuehn, V. Suessle, M.J. Burnett, C.T. Downs, A. Weinmann, and E. Hergenroether. Automated bioacoustic monitoring for south african bird species on unlabeled data. *arXiv preprint arXiv:2406.13579*, 2024.

Derek Engelbrecht. Vocalizations and song flight of pink-billed lark spizocorys conirostris. *Afrotropical Bird Biology: Journal of the Natural History of African Birds*, 1, 2021.

B. Ghani, T. Denton, S. Kahl, and H. Klinck. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports*, 13(1):22876, 2023.

M. Hagiwara. Introducing birdaves: Self-supervised audio foundation model for birds. `https://www.earthspecies.org/blog/introducing-birdaves-self-supervised-audio-foundation-model-for-birds`, 2024. [Accessed: 14 November 2024].

Jan Huus, Kevin G Kelly, Erin M Bayne, and Elly C Knight. Hawkears: A regional, high-performance avian acoustic classifier, 2024.

S. Kahl. Re: Birdnet-analyzer discussion topic #234. `https://github.com/kahst/BirdNET-Analyzer/discussions/234`, January 2024. [Accessed: 14 November 2024].

S. Kahl, C.M. Wood, M. Eibl, and H. Klinck. Birdnet: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61:101236, 2021.

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*, 13(12):2799–2810, 2022.

Alan TK Lee, Michael Brooks, and Les G Underhill. The sabap2 legacy: A review of the history and use of data generated by a long-running citizen science project. *South African Journal of Science*, 118(1-2):1–4, 2022.

T. Mosikidi, N. Le Maitre, S.L. Steenhuisen, V.R. Clark, K.S. Lloyd, and A. Le Roux. Passive acoustic monitoring detects new records of globally threatened birds in a high-elevation wetland (free state, south africa). *Bird Conservation International*, 33:e80, 2023.

I. Moummad, N. Farrugia, and R. Serizel. Self-supervised learning for few-shot bird sound classification. In *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pages 600–604. IEEE, April 2024.

H. Nel, A.K. Mishra, and F. Schonken. Ecosonicml: Harnessing machine learning for biodiversity monitoring in south african wetlands. *SN Computer Science*, 5(5):1–9, 2024.

C. Pérez-Granados. Birdnet: applications, performance, pitfalls and future opportunities. *Ibis*, 165(3):1068–1075, 2023.

Brian L Sullivan, Christopher L Wood, Marshall J Iliff, Rick E Bonney, Daniel Fink, and Steve Kelling. ebird: A citizen-based bird observation network in the biological sciences. *Biological conservation*, 142(10):2282–2292, 2009.

B. Van Merriënboer, J. Hamer, V. Dumoulin, E. Triantafillou, and T. Denton. Birds, bats and beyond: Evaluating generalization in bioacoustics models. *Frontiers in Bird Science*, 3:1369756, 2024.

Y. Yang, K. Zhou, N. Trigoni, and A. Markham. Ssl-net: A synergistic spectral and learning-based network for efficient bird sound classification. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 926–930. IEEE, April 2024.