

Active Sampling for Data Efficient Large-Scale Bioacoustic Monitoring

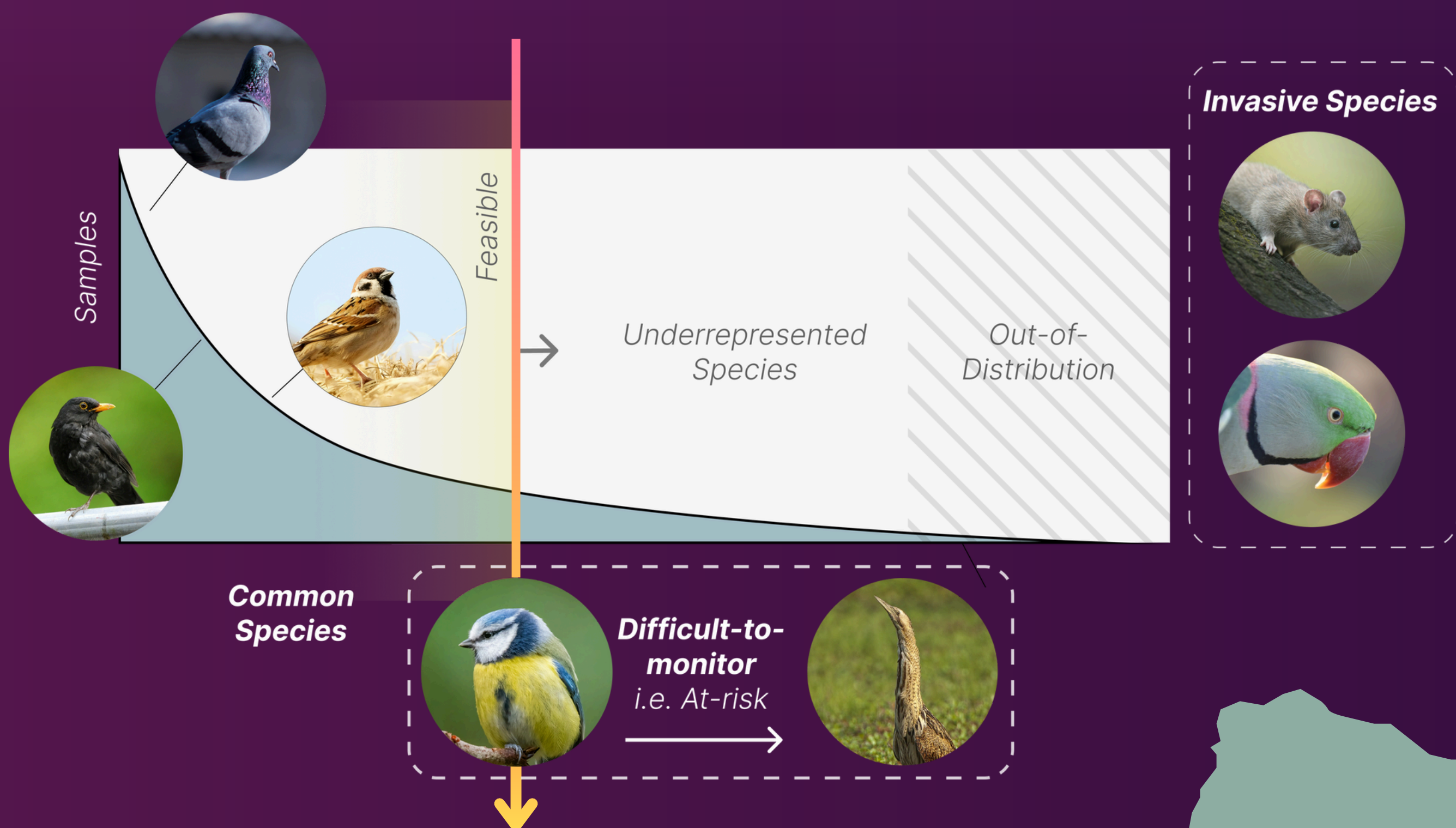


Introduction

The sounds of nature provide critical biodiversity insights. Recent advances in computational bioacoustics have transformed biodiversity monitoring but reliable model inference at a transnational-scale remains a challenge.

Challenges

In large-scale deployments, it is typical for a few common species to be overrepresented, while at-risk species are underrepresented (long-tail distribution). Performance of AI classification tends to degrade along this tail, making **reliable detection of underrepresented species** especially challenging. This challenge is complicated by **spatiotemporal variation, scale** and **limited annotation resources**.

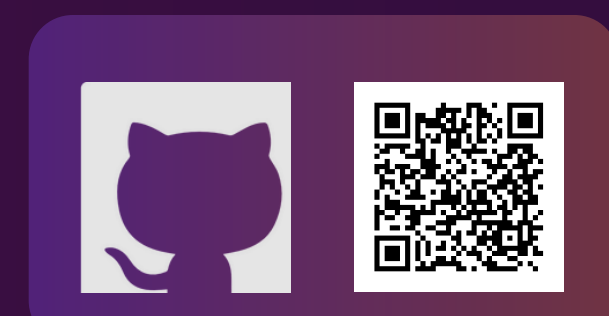


Objectives

- Develop scalable data pipeline for transnational biodiversity monitoring
- Investigate data-efficient expert integration with active learning

Results

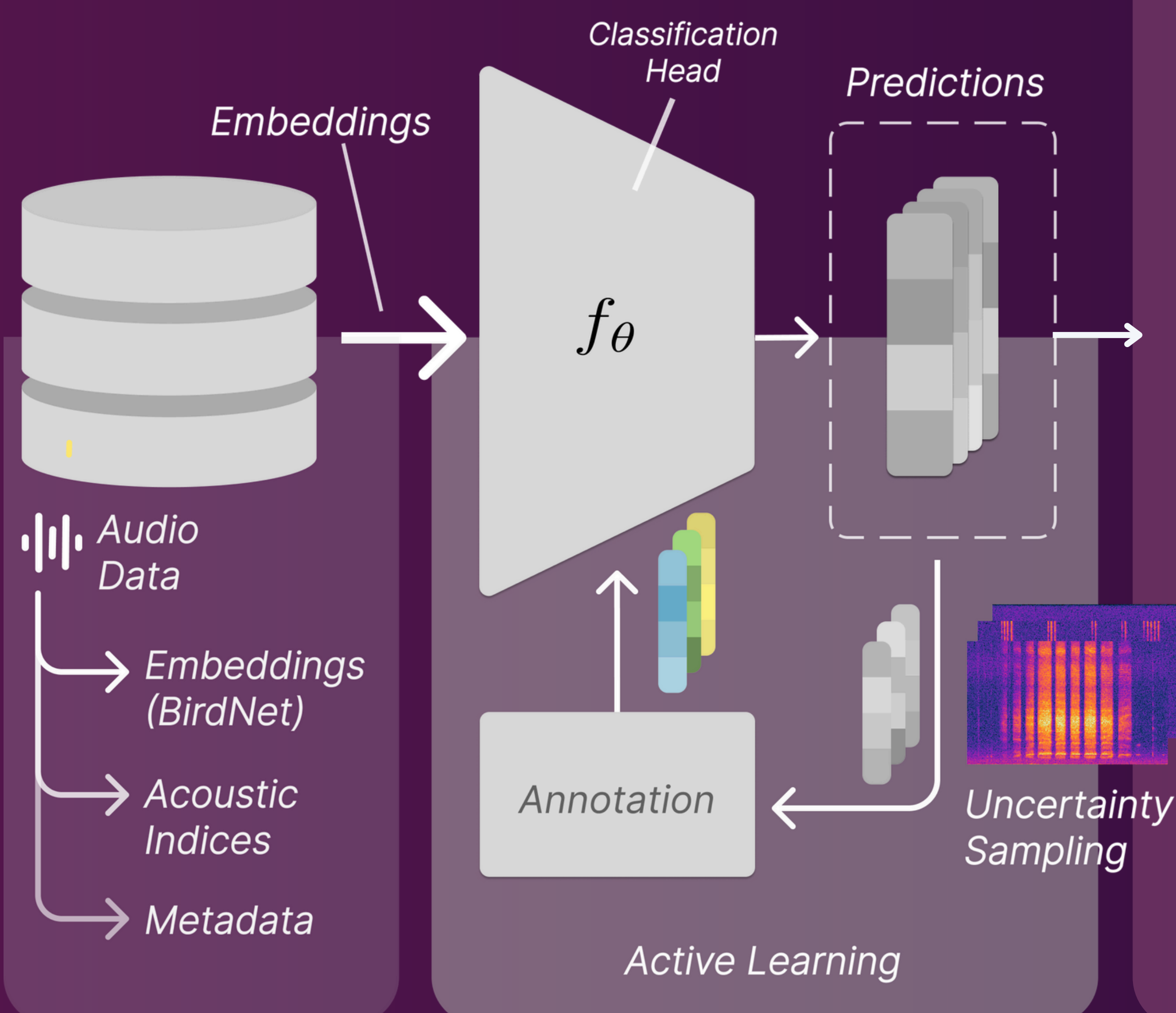
Active learning enables more reliable and scalable bioacoustic monitoring by reducing labeling requirements (training efficiency) and improving species detection especially for underrepresented species. Initial results using WABAD dataset.



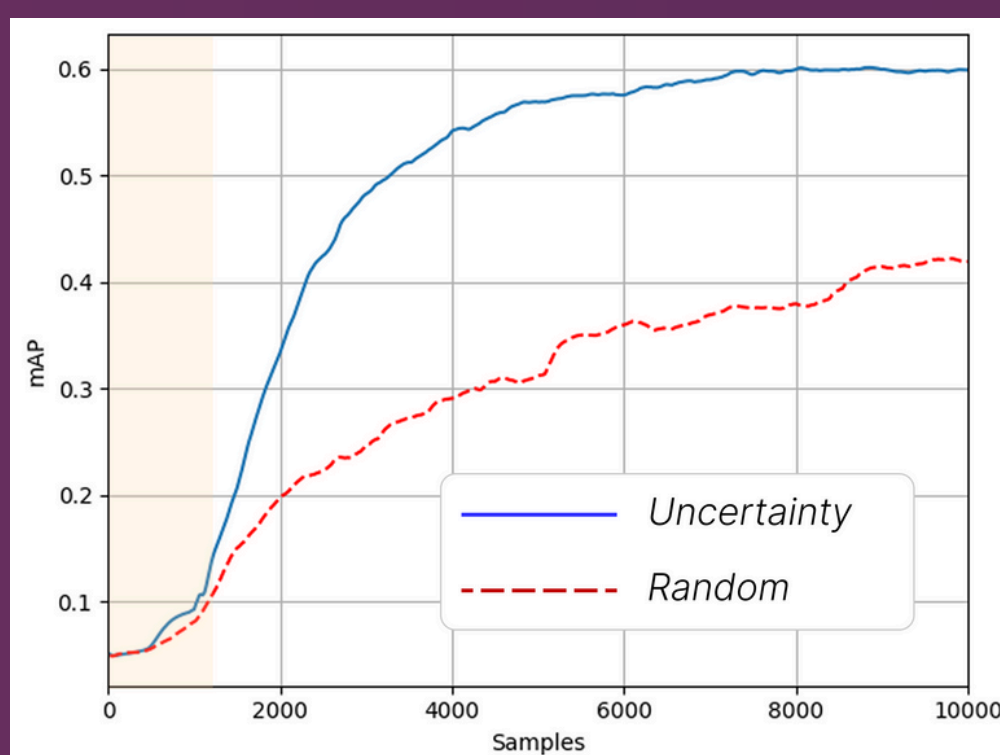
Access to data pipeline

Building a Data Pipeline

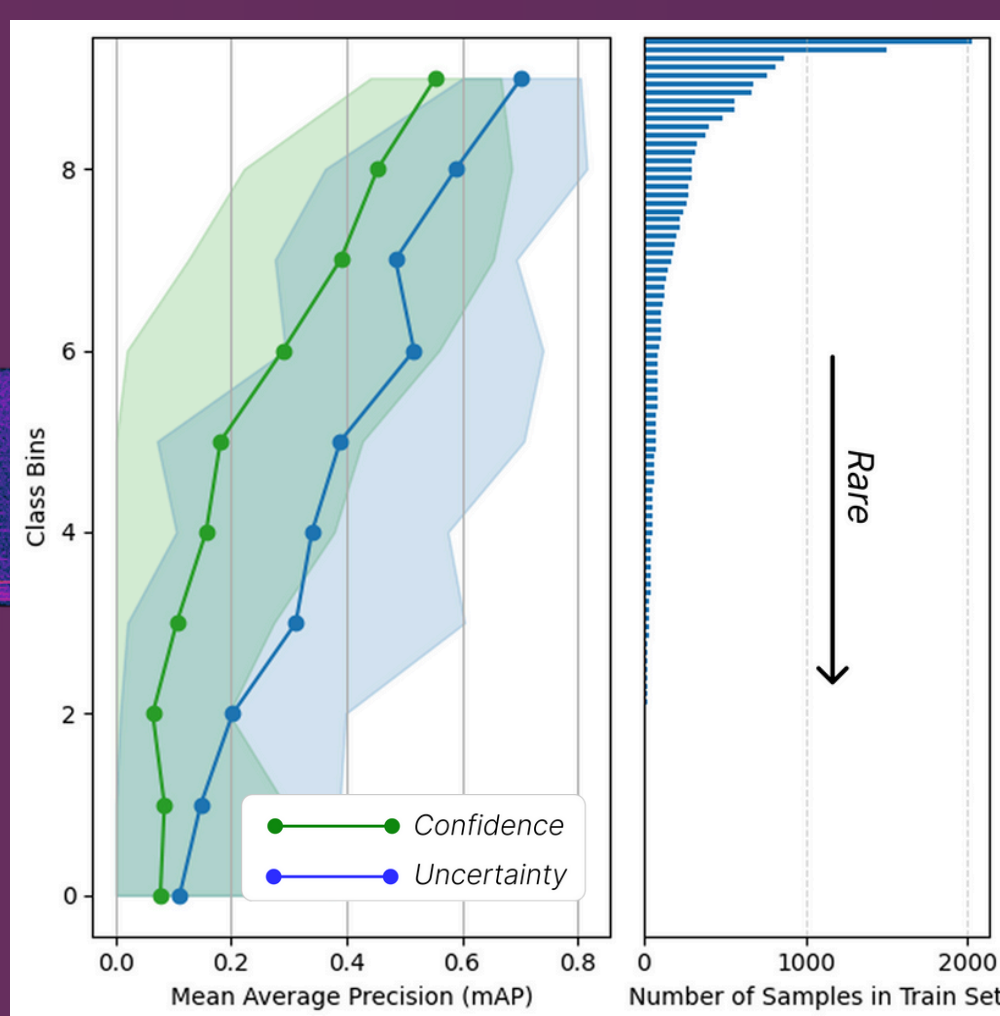
We have built a data pipeline which processes continuous 24/7 acoustic data from up to 150 devices from the top of Norway to Spain.



Improving Training Efficiency



Improving Performance for Underrepresented Species



Collaborators

Consortium Members

Corentin Bernard², Benjamin Cretois³, Cynthia Barile⁵, Julia Wiel³, Dan Stowell^{1,4}, Eva Serrano-Davies⁶, Carolyn Rosten³, Daniel Kissling⁵, Ricard Marxer², Hervé Glotin², Dani Villero⁶, Lluís Brotons⁶, Cristian Pérez-Granados⁶, Gerard Bota⁶, Sarab Sethi⁷

Affiliations

¹Tilburg University, ²Toulon University, ³Norwegian Institute for Nature Research (NINA), ⁴Naturalis Biodiversity Center, ⁵University of Amsterdam, ⁶Forest Science Center of Catalonia, ⁷Imperial College London

Funded by **Biodiversa+**

Presented by:



Dr. Ben McEwen¹

Postdoctoral Researcher
b.j.mcewen@tilburguniversity.edu
Audio Machine Learning and Bioacoustics
Co-lead of the Listening Lab