Demo: Enhancing Wildlife Acoustic Data Annotation Efficiency through Transfer and Active Learning

Hannes Kath[1,2], **Patricia P. Serafini**[3,4], **Ivan B. Campos**[3,5], **Thiago S. Gouvêa**[1,2], **Daniel Sonntag**[1,2]
[1]Interactive Machine Leraning, German Research Center for Artificial Intelligence (DFKI), Germany
[2]Applied Artificial Intelligence, Carl von Ossietzky University of Oldenburg, Germany
[3]National Center for Wild Bird Conservation and Research (CEMAVE), Chico Mendes Institute for Biodiversity Conservation (ICMBio), Brazil
[4]Universidade Federal de Santa Catarina (UFSC), Brazil
[5]Departamento de Biologia Geral, Universidade Federal de Minas Gerais (UFMG), Brazil
{hannes.kath, thiago.gouvea, daniel.sonntag}@dfki.de
{patricia.serafini, ivan.campos}@icmbio.gov.br

Abstract

Passive Acoustic Monitoring (PAM) has become a key technology in wildlife monitoring, generating large amounts of acoustic data. However, the effective application of machine learning methods for sound event detection in PAM datasets is highly dependent on the accessibility of annotated data, a process that can be labour intensive. As a team of domain experts and machine learning researchers, in this paper we present a no-code annotation tool designed for PAM datasets that incorporates transfer learning and active learning strategies to address the data annotation challenge inherent in PAM. Transfer learning is applied to use pretrained models to compute meaningful embeddings from the PAM audio files. Active learning iteratively identifies the most informative samples and then presents them to the user for annotation. This iterative approach improves the performance of the model compared to random sample selection. In a preliminary evaluation of the tool, a domain expert annotated part of a real PAM data set. Compared to conventional tools, the workflow of the proposed tool showed a speed improvement of 2-4 times. Further enhancements, such as the incorporation of sound examples, have the potential to further improve efficiency.

1 Introduction

Biodiversity loss is among the most pressing issues of our days [Cardinale *et al.*, 2012]. Drivers of the negative change have been accelerating, and meeting internationally agreed conservation targets will require transformative change [IPBES, 2019]. While machine learning methods have been increasingly brought to bear to support wildlife management, the tools available to those on the ecological front lines still lag the state-of-the-art in artificial intelligence research [Tuia *et al.*, 2022; Gouvêa *et al.*, 2023]. Passive acoustic monitoring (PAM) has emerged as a powerful technology for wildlife monitoring, allowing ecologists to gather extensive data on wildlife with minimal disturbance of habitats [Sugai *et al.*, 2019; Sugai and Llusia, 2019]. PAM systems can be used to continuously record sounds from various biomes, offering valuable insights into animal behaviour, species richness, and ecosystem health, with applications in ecosystem management, rapid biodiversity assessments [Sueur *et al.*, 2008], and basic research [Ross *et al.*, 2023].

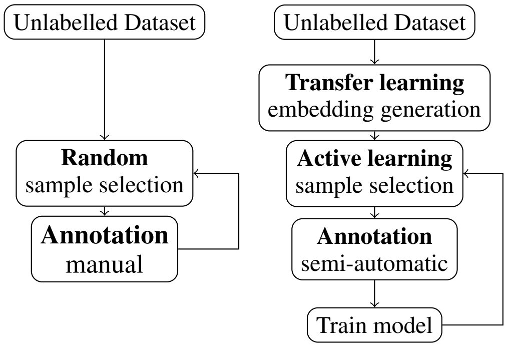
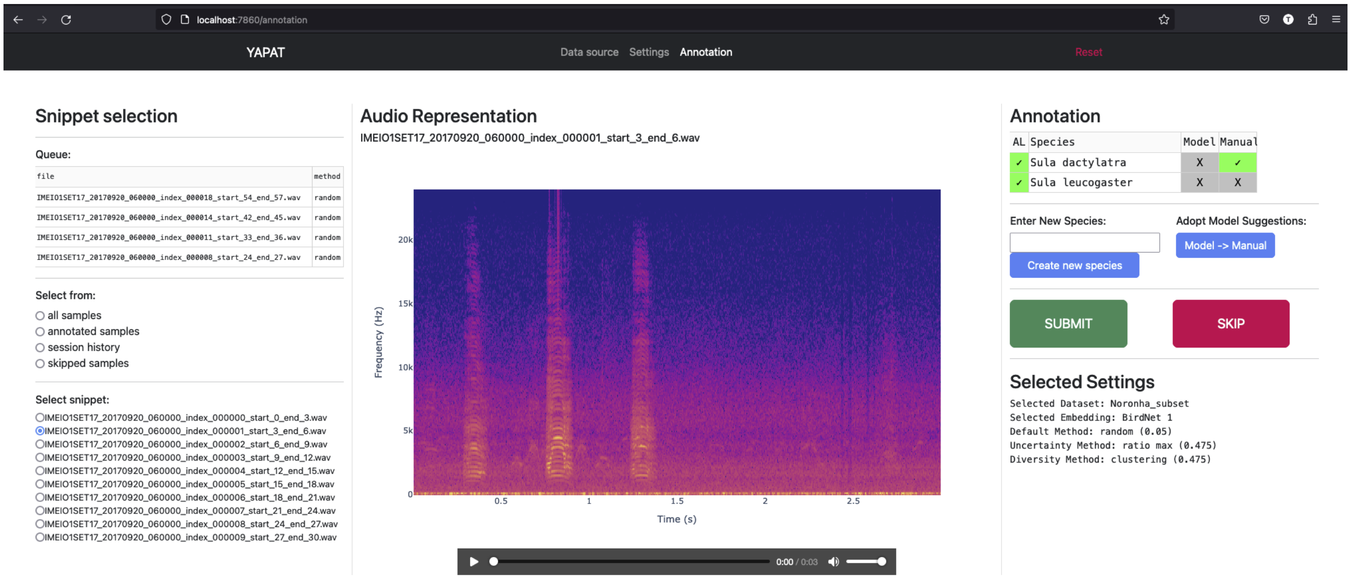


Figure 1: Workflow for annotating PAM datasets, comparing the conventional approach (left) with the proposed approach (right).

However, effectively utilising this vast amount of data still poses significant management and analysis challenges. Sound event detection in particular (e.g., species identification) is limited by the need for annotated data to train supervised machine learning models. PAM data annotation is usually done manually, in a laborious and time-consuming process: domain experts listen to each audio file, annotating events by manually selecting time segments on a graphical representation of the sound (e.g., amplitude envelope or spectrogram) [Tkachenko *et al.*, 2020; Perry *et al.*, 2021; Cañas *et al.*, 2023]. This approach is laborious and incompatible with the large volume of data generated by PAM, and current research focuses on automating the annotation process. Seadash introduces a graphical implementation of data programming, but lacks evaluation on real life datasets



Snippet selection

Queue:

file	method
IMEIO1SET17_20170920_060000_index_000018_start_54_end_57.wav	random
IMEIO1SET17_20170920_060000_index_000014_start_42_end_45.wav	random
IMEIO1SET17_20170920_060000_index_000011_start_33_end_36.wav	random
IMEIO1SET17_20170920_060000_index_000008_start_24_end_27.wav	random

Select from:

○ all samples
○ annotated samples
○ session history
○ skipped samples

Select snippet:

○IMEIO1SET17_20170920_060000_index_000000_start_0_end_3.wav
●IMEIO1SET17_20170920_060000_index_000001_start_3_end_6.wav
○IMEIO1SET17_20170920_060000_index_000002_start_6_end_9.wav
○IMEIO1SET17_20170920_060000_index_000003_start_9_end_12.wav
○IMEIO1SET17_20170920_060000_index_000004_start_12_end_15.wav
○IMEIO1SET17_20170920_060000_index_000005_start_15_end_18.wav
○IMEIO1SET17_20170920_060000_index_000006_start_18_end_21.wav
○IMEIO1SET17_20170920_060000_index_000007_start_21_end_24.wav
○IMEIO1SET17_20170920_060000_index_000008_start_24_end_27.wav
○IMEIO1SET17_20170920_060000_index_000009_start_27_end_30.wav

Audio Representation

IMEIO1SET17_20170920_060000_index_000001_start_3_end_6.wav

Annotation

AL	Species	Model	Manual
✓	Sula dactylatra	X	✓
✓	Sula leucogaster	X	X

Enter New Species:

Adopt Model Suggestions:

Selected Settings

```
Selected Dataset: Noronha_subset
Selected Embedding: BirdNet 1
Default Method: random (0.05)
Uncertainty Method: ratio max (0.475)
Diversity Method: clustering (0.475)
```

Figure 2: Layout of the annotation interface. **(left)** Sample 3-s snippets, embedded with a user-selected transfer-learning model, are selected by multi-class active learning algorithms and queued for annotation. **(center)** The selected snippet is represented as a spectrogram and can be played back.**(right)** Labels (e.g., species id) are assigned to the selected sample by ticking a box on the *manual* column of the list of classes, or by accepting model suggestions. The *AL* column selects classes for which active learning optimises sample selection.

[Gouvêa *et al.*, 2022]. DetEdit allows simultaneous detection of bouts of events through a configurable signal processing pipeline that includes a GUI for accepting/rejecting detections; it runs on a proprietary platform, and has only been evaluated on odontocete echolocation click datasets [Solsona-Berga *et al.*, 2020]. Scikit-maad [Ulloa *et al.*, 2021] and Bam-Scape [Michaud *et al.*, 2022] are tools for large scale PAM data analysis by spectrogram segmentation and clustering; as command line tools, they lack interactivity.

A promising, under-explored direction of research on interactive PAM annotation is the leveraging of small amounts of available labels (e.g., at early stages of annotation of a novel dataset) to improve the efficiency of the annotation process. We have previously shown that early annotations can be used by a semi-supervised dimensionality reduction method to generate actionable scatter-plots, speeding up the annotation process by facilitating grouping of similar samples as well as identification of outliers [Kath *et al.*, 2023].

As an interdisciplinary team of ecologists, ecosystems managers, and machine learning researchers, in the present work we introduce a PAM annotation tool that builds on our previous work by leveraging active learning [Kadir *et al.*, 2023] in an embedding space learned in a related domain.

2 System Description

Machine Learning Backend. It has been shown that BirdNet, a neural model trained primarily on focal recordings of songbirds [Kahl *et al.*, 2021], learns embeddings that are useful for sound event detection on PAM data of different species groups [Ghani *et al.*, 2023; Kath *et al.*, 2024a]. Following [Kath *et al.*, 2024a], our system splits sound inputs into 3-s segments and embeds them into one of the three penultimate layers of BirdNet. A suite of sampling algorithms for multi-label active learning are subsequently used to select segments for annotation. The suite of methods includes uncertainty-based (average entropy, least-confidence, or ratio, as well as max ratio) and diversity-based methods (see [Kath *et al.*, 2024a] for details.) In the tool introduced here, the user makes use of a graphical interface to control the choice of embedding layer and active learning sampling methods, as well as for evaluating and annotating each sample.

User Interface. We designed a multi-tab Dash[1] web application composed of three interfaces, namely the Data Source tab, the Settings tab and the Annotation tab[2] (see fig. 2). The Data Source tab controls whether the input data is read from the local computer or a remote location—in the current version we considered practical to make use of a commercial cloud provider (Google Drive). The settings tab controls the choice of transfer learning embedding model and layer, and the composition of active learning sampling methods. In addition, data preprocessing is triggered from this tab: each file in the selected dataset is split into 3-second audio segments (snippets), and their corresponding embeddings are computed using the selected embedding model and layer. Lastly, the Annotation tab presents the selected snippets along with resources for evaluation and label assignment. The layout of this tab is composed of three columns: snippet selection, audio representation, and annotation (fig. 2). The snippet selection area lists the sound snippets queued for annotation, along

[1] https://dash.plotly.com/

[2] https://youtu.be/mAVm79-UmlA

with the sampling method used to select them. In addition, it is also possible to select samples manually. The Audio Representation area displays the currently selected sample as a spectrogram. Selecting a time-frequency box on the spectrogram allows listening to the corresponding band-pass filtered segment. The Annotation area allows assigning labels (e.g., species id) to the selected audio snippet by ticking boxes on the list of label classes. New classes can be entered through a text input field. The active learning model can be used for inference as well: predicted labels can be accepted by clicking the *Model→Manual* button. A text display on the bottom-right corner indicates settings chosen on the Settings tab.

Workflow. While fig. 1 illustrates the data-centric workflow, here we outline the workflow from the user perspective. When the web application is accessed, only the first tab (data source) is available. Upon selection of a data location, the system validates it (either by checking the local path or an authorisation code provided by Google Drive.) Next, on the Settings tab the user can select a dataset folder, as well as an embedding model and layer, and an active learning strategy. The active learning strategy is composed of controllable fractions of random sampling (as in ϵ-greedy policies), a diversity-based method, and a choice of uncertainty-based method (see [Kath *et al.*, 2024a] for details.) This composite strategy is used to select the audio samples (snippets) to be presented to the user. Next, the user triggers the pre-processing of the data, a process cached separately for different transfer-learning settings values. While these settings remain modifiable at any time, the actual annotation of data takes place in the Annotation tab, which becomes active after pre-processing. The user determines the species present in the selected sample by examining the spectrogram and listening to the audio. By selecting a time-frequency region in the spectrogram, the audio is band-pass filtered, in some cases allowing the user to identify species calls even when obscured by loud background noise (provided they don't overlap in time and/or frequency). The user selects the present species in the Manual column of the annotation table. If the active learning model provides satisfactory predictions the user can simply accept them by pressing the *Model→Manual* button. Active learning gains deteriorate when optimising for too many classes at the same time [Kath *et al.*, 2024b]; the tool therefore allows the user to mark to which label classes should active learning attend by ticking the boxes on the AL column of the annotation pane (fig. 2). After annotating the sample snippet, the user saves the annotations via the Submit button. If the user is unsure of the species present, they can skip the annotation for later review using the Skip button. Pressing the Submit or Skip button loads the next sample for annotation. If the user wishes to select a sample manually, e.g. to inspect snippets based on metadata, or to revisit skipped snippets, they can do so using the Snippet Selection area.

3 Preliminary Evaluation

Preliminary evaluations are being conducted using a real life dataset collected at Fernando de Noronha, a designated protected area recognised by UNESCO as a World Natural Heritage site. The application was deployed on HuggingFace where a domain expert (who co-authors this study) used the tool for identifying and labelling vocalisations by oceanic birds. The domain expert described the workflow as particularly intuitive, although opportunities for improvement were identified. In particular, when there is uncertainty about the species identified, it is common practice to listen to example files within the Xeno-Canto database[3]. The integration of example links within the tool is suggested to streamline and improve the workflow.

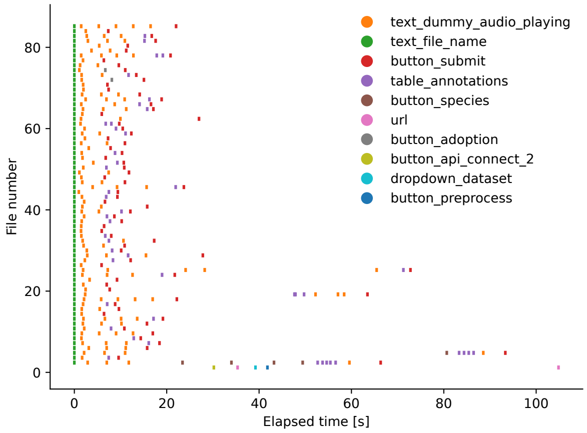


Figure 3: Example session. Rows are annotations of single 3-second segments, and ticks are interaction events.

In addition, every interaction with GUI elements is being logged for a subsequent usability study. Figure 3 shows raw results from an example annotation session. A previous study with this same dataset and user [Lüers *et al.*, 2024] showed that, using the traditional tools, the domain expert needs 648 ± 301 seconds to annotate each 60-seconds long input file, a factor of 10.8 ± 5.0 relative to real time (mean $\pm$ standard deviation). In the annotation session depicted in fig. 3, the median annotation time for a 3-second audio file was 11 seconds with our tool, a factor of less than 4. While anecdotal, this represents a promising first piece of evidence that our tool can have a significant impact on the manageability of large PAM datasets.

4 Conclusion and Future Work

We present an annotation tool designed for passive acoustic monitoring datasets that uses a combination of transfer learning and active learning strategies. This approach aims to generate meaningful representations of the data and present the most informative samples to the user for annotation. The tool includes an interactive user interface designed to be applicable to real-world problems. Preliminary evaluation shows promising results, with a remarkable acceleration of the conventional annotation process by a factor of 2-4, accompanied by an intuitive user experience. Future work includes conducting a more extensive user study and incorporating requested features. One notable enhancement is the integration of sound samples from sources such as Xeno Canto.

[3] https://xeno-canto.org/

Acknowledgements

This research is part of the Computational Sustainability & Technology project area[4], and has been supported by the Ministry for Science and Culture of Lower Saxony (MWK), the Endowed Chair of Applied Artificial Intelligence, Oldenburg University, and DFKI.

References

[Cardinale *et al.*, 2012] Bradley J. Cardinale, J. Emmett Duffy, Andrew Gonzalez, David U Hooper, Charles Perrings, Patrick A. Venail, Anita Narwani, Georgina M. Mace, David Tilman, David A. Wardle, Ann Kinzig, Gretchen C. Daily, Michel Loreau, J. Grace, Anne Larigauderie, Diane S. Srivastava, and Shahid Naeem. Biodiversity loss and its impact on humanity. *Nature*, 486:59–67, 2012.

[Cañas *et al.*, 2023] Juan Cañas, María Toro-Gómez, Larissa Sugai, et al. A dataset for benchmarking neotropical anuran calls identification in passive acoustic monitoring. *Scientific Data*, 10(1):771, 2023.

[Ghani *et al.*, 2023] Burooj Ghani, Tom Denton, Stefan Kahl, and Holger Klinck. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. *Scientific Reports*, 13(1):22876, 2023.

[Gouvêa *et al.*, 2022] Thiago S. Gouvêa, Ilira Troshani, Marc Herrlich, and Daniel Sonntag. Annotating sound events through interactive design of interpretable features. In Stefan Schlobach, María Pérez-Ortiz, and Myrthe Tielman, editors, *HHAI 2022: Augmenting Human Intellect - Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence, Amsterdam, The Netherlands, 13-17 June 2022*, volume 354 of *Frontiers in Artificial Intelligence and Applications*, pages 305–306. IOS Press, 2022.

[Gouvêa *et al.*, 2023] Thiago Gouvêa, Hannes Kath, Ilira Troshani, et al. Interactive Machine Learning Solutions for Acoustic Monitoring of Animal Wildlife in Biosphere Reserves. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 6405–6413, Macau, SAR China, 2023.

[IPBES, 2019] IPBES. Summary for policymakers of the global assessment report on biodiversity and ecosystem services. Technical report, Zenodo, November 2019. Version Number: summary for policy makers.

[Kadir *et al.*, 2023] Abdul Kadir, Hasan Alam, and Daniel Sonntag. Edgeal: An edge estimation based active learning approach for oct segmentation. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, pages 79–89, 2023.

[Kahl *et al.*, 2021] Stefan Kahl, Connor Wood, Maximilian Eibl, and Holger Klinck. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61:101236, 2021.

[Kath *et al.*, 2023] Hannes Kath, Thiago Gouvêa, and Daniel Sonntag. A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 7140–7144, Macau, SAR China, 2023.

[Kath *et al.*, 2024a] Hannes Kath, Patricia P. Serafini, Ivan Braga Campos, Thiago S Gouvêa, and Daniel Sonntag. Leveraging Transfer Learning and Active Learning for Sound Event Detection in Passive Acoustic Monitoring of Wildlife. In *3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*, Vancouver, Canada, 2024.

[Kath *et al.*, 2024b] Hannes Kath, Thiago S Gouvêa, and Daniel Sonntag. Active learning in multi-label classification of bioacoustic data. Under review, 2024.

[Lüers *et al.*, 2024] Bengt Lüers, Patricia P. Serafini, Ivan Braga Campos, Thiago S Gouvêa, and Daniel Sonntag. BirdNET-Annotator: AI-Assisted Strong Labelling of Bird Sound Datasets. In *3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE)*, Vancouver, Canada, 2024.

[Michaud *et al.*, 2022] Félix Michaud, Jérôme Sueur, Maxime LE Cesne, and Sylvain Haupert. Unsupervised classification to improve the quality of a bird song recording dataset. *ArXiv*, abs/2302.07560, 2022.

[Perry *et al.*, 2021] Sean Perry, Vaibhav Tiwari, Nishant Balaji, Erika Joun, Jacob Ayers, Mathias Tobler, Ian Ingram, Ryan Kastner, and Curt Schurgers. Pyrenote: a web-based, manual annotation tool for passive acoustic monitoring. In *IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems, MASS 2021, Denver, CO, USA, October 4-7, 2021*, pages 633–638. IEEE, 2021.

[Ross *et al.*, 2023] Samuel Ross, Darren O'Connell, Jessica Deichmann, et al. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. *Functional Ecology*, 37(4):959–975, 2023.

[Solsona-Berga *et al.*, 2020] Alba Solsona-Berga, Kaitlin E. Frasier, Simone Baumann-Pickering, Sean M. Wiggins, and John A. Hildebrand. Detedit: A graphical user interface for annotating and editing events detected in long-term acoustic monitoring data. *PLoS Comput. Biol.*, 16(1), 2020.

[Sueur *et al.*, 2008] Jérôme Sueur, Sandrine Pavoine, Olivier Hamerlynck, and Stéphanie Duvail. Rapid Acoustic Survey for Biodiversity Appraisal. *PLOS ONE*, 3(12):e4065, 2008.

[Sugai and Llusia, 2019] Larissa Sugai and Diego Llusia. Bioacoustic time capsules: Using acoustic monitoring to document biodiversity. *Ecological Indicators*, 99:149–152, 2019.

[Sugai *et al.*, 2019] Larissa Sugai, Thiago Silva, José Ribeiro, and Diego Llusia. Terrestrial Passive Acoustic Monitoring: Review and Perspectives. *BioScience*, 69(1):15–25, 2019.

[4] https://cst.dfki.de/

[Tkachenko *et al.*, 2020] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. Label Studio: Data labeling software, 2020.

[Tuia *et al.*, 2022] Devis Tuia, Benjamin Kellenberger, Sara Beery, Blair R. Costelloe, Silvia Zuffi, Benjamin Risse, Alexander Mathis, Mackenzie W. Mathis, Frank van Langevelde, Tilo Burghardt, Roland Kays, Holger Klinck, Martin Wikelski, Iain D. Couzin, Grant van Horn, Margaret C. Crofoot, Charles V. Stewart, and Tanya Berger-Wolf. Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1):792, February 2022. Number: 1 Publisher: Nature Publishing Group.

[Ulloa *et al.*, 2021] Juan Sebastián Ulloa, Sylvain Haupert, Juan Felipe Latorre, Thierry Aubin, and Jérôme Sueur. scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. *Methods in Ecology and Evolution*, pages 2041–210X.13711, September 2021.