Enhancing Biodiversity Monitoring: An Interactive Tool for Efficient Identification of Species in Large Bioacoustics Datasets

Hannes Kath
University of Oldenburg
Oldenburg, Germany
German Research Center for Artificial Intelligence (DFKI)
Oldenburg, Germany
hannes.kath@dfki.de

Ilira Troshani
University of Oldenburg
Oldenburg, Germany
German Research Center for Artificial Intelligence (DFKI)
Oldenburg, Germany
ilira.troshani@dfki.de

Bengt Lüers
German Research Center for Artificial Intelligence (DFKI)
Oldenburg, Germany
bengt.lueers@dfki.de

Thiago S. Gouvêa
German Research Center for Artificial Intelligence (DFKI)
Oldenburg, Germany
thiago.gouvea@dfki.de

Daniel Sonntag
University of Oldenburg
Oldenburg, Germany
German Research Center for Artificial Intelligence (DFKI)
Saarbrücken, Germany
daniel.sonntag@dfki.de

ABSTRACT

Biodiversity loss is a major challenge for humanity, which has increased the rate of species extinction by a factor of 100-1000 compared to pre-industrial times. XPRIZE Rainforest is a competition focused on developing a pipeline for real-time biodiversity measurement: teams have 24 hours to collect data and another 48 hours to produce a list of species present in the data. Passive acoustic monitoring (PAM) is a scalable technology for data acquisition in wildlife monitoring. However, analyzing large PAM datasets poses a significant challenge. This paper presents a tool used by the Brazilian team during the XPRIZE Rainforest finals. Using a combination of audio separation, weakly supervised learning, transfer learning, active learning, multiple-instance learning, and novel class detection, samples are carefully selected and presented to the user for annotation.

CCS CONCEPTS

• **Human-centered computing**; • **Computing methodologies** → **Machine learning algorithms**; *Search methodologies*;

KEYWORDS

passive acoustic monitoring, novel class detection, transfer learning, active learning

ACM Reference Format:

Hannes Kath, Ilira Troshani, Bengt Lüers, Thiago S. Gouvêa, and Daniel Sonntag. 2024. Enhancing Biodiversity Monitoring: An Interactive Tool for Efficient Identification of Species in Large Bioacoustics Datasets. In *Proceedings of 26th ACM International Conference on Multimodal Interaction (ICMI '24)*. ACM, New York, NY, USA, 3 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Biodiversity. 'The variability among living organisms from all sources including, inter alia, terrestrial, marine, and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species, and of ecosystems' is the definition of biodiversity from the Convention on Biological Diversity (CBD) [25]. Biodiversity, in simplest terms, refers to the variety of life. The loss of biodiversity is among the most serious issues of our days [1]. Human activities have accelerated the extinction rate to 100-1000 times higher than pre-industrial levels [3, 15], necessitating transformative change to achieve international conservation goals [6].

XPRIZE Rainforest. The urgency of biodiversity loss has led to incentives to automate scalable biodiversity measurement. XPRIZE Rainforest 'is a global 5-year, $10 million competition that convenes innovators and experts across disciplines [...] and challenges them to use novel technologies to expedite the monitoring of tropical biodiversity'[1]. After a 24-hour survey of 100 hectares of tropical rainforest, the data is analyzed within 48 hours to generate a list of species present. Methods are needed that maximize species identification in an unlabelled dataset while minimizing the number of samples examined, as formally stated in [9].

Passive acoustic monitoring (PAM). Monitoring the impacts of human activities requires scalable methods for effective biodiversity measurement. PAM is a scalable, non-invasive technology designed for wildlife monitoring, enabling the collection of extensive data with minimal habitat disturbance [19, 20]. PAM systems can be used to provide continuous sound recordings across different

[1] https://www.xprize.org/prizes/rainforest

ICMI '24, November 04–08, 2024, San José, Costa Rica

ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX

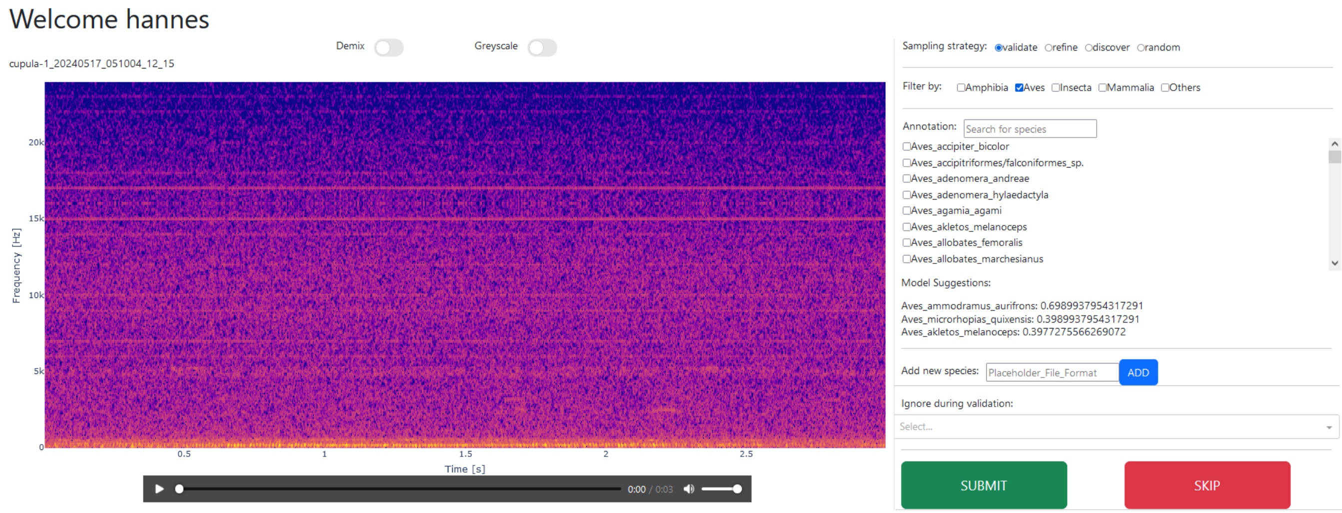

Figure 1: Layout of the user interface. (left) The selected sample is presented with the file name, spectrogram, and playback option. (right) Annotation column featuring selection panels for the sampling strategy, competence class, and species list.

biomes, providing valuable insights into animal behaviour, species richness and ecosystem health. They are used for ecosystem management, rapid biodiversity assessments [18] and basic research [16]. While data acquisition for wildlife monitoring is straightforward and cost-effective, evaluating large-scale audio datasets presents significant challenges. Machine learning is increasingly employed to address this challenge, but existing tools still lag behind state-of-the-art artificial intelligence research [5, 23].

PAM annotation tools. Predominant annotation tools for PAM data rely on a laborious and time-consuming manual process: domain experts select files, listen to the audio, review graphical sound representations (e.g., amplitude envelope or spectrogram), and annotate events within these representations [2, 14, 21]. As this approach is not scalable, current research is focused on automating the annotation process. Seadash [4] uses data programming techniques but lacks evaluation on real-world datasets. DetEdit [17] accelerates annotation by enabling simultaneous detection of event bouts, but it operates on a proprietary platform and has only been evaluated on odontocete echolocation click datasets. Scikit-maad [24] and BamScape [13] are designed for large-scale PAM data analysis through spectrogram segmentation and clustering. However, as command-line tools, they lack accessibility and interactivity. An interactive clustering-based annotation tool is presented in [8], but it has not been evaluated on real data. BirdNET-Annotator [12] uses BirdNet [7], a neural model trained on vocal bird recordings, to annotate samples with likely contained bird species. Using BirdNet as embedding model and applying active learning methods for sample selection achieves superior results across various PAM datasets [11] and extends BirdNET-Annotator to non-avian species, with promising outcomes demonstrated in an initial user study [10].

Contribution. This paper presents an annotation tool used by the Brazilian team in the final of the XPRIZE Rainforest competition. Using audio separation, weakly supervised learning, transfer learning, active learning, multiple-instance learning and novel class detection methods, audio samples are carefully selected and presented within our user interface for annotation[2].

2 SYSTEM DESCRIPTION

As a pre-processing step, all incoming audio is resampled to 48 kHz and split into 3 s chunks before a sound separation model, pre-trained on Amazonian recordings, is used to create 8 channels from each chunk.

2.1 Sampling strategies

The sampling strategies 'refine' and 'discover' input sample embeddings generated by a variational autoencoder. The multiple-instance learning models trained for these strategies use all annotated normalized samples and their corresponding sound separations.

validate. A model, pre-trained in a weakly supervised manner on large sound collections [22] such as Xeno-Canto[3] for all relevant and present species, is applied to the normalized audio and the separations. For each sample, the model outputs the likelihood of each species being present. Samples with a high likelihood of containing a species that is not yet annotated at least 5 times are selected for user validation.

refine. A model is trained to predict the probability of each species being present in a given sample. After applying the model to the unlabelled data, the active learning uncertainty sampling strategy 'ratio' with the best-performing score aggregation method 'max' [11] selects samples to refine the model.

discover. One model is trained to predict the probability of each species being present in a given sample, and another to predict whether any species is present in a given sample. To discover new species, samples likely to contain any species but unlikely to contain a species already found are selected.

[2]https://youtu.be/2dNYQvxn7cE
[3]https://xeno-canto.org/

random. Samples are randomly selected.

2.2 User interface

Figure 1 shows the user interface, a two-column web application.

The left column displays the selected sample. At the top, the file name indicating the day of the year and time of day is displayed. In the center, a spectrogram visualizes the selected audio file, and the mouse can zoom in on regions of interest. The 'Demix' switch changes the display to show the concatenation of the original sample and the 8 channels generated by the sound separation model. The 'Greyscale' switch changes the colourmap to greyscale. The bottom of the column displays the auditory representation, aligned with the visible spectrogram. Zooming into the spectrogram filters the audio in time and frequency, and activating the 'Demix' switch changes the audio to the concatenation of the original sample and the 8 channels generated by the sound separation model.

The right column displays all the features required for the annotation process. At the very top, the user selects a sampling strategy (see section 2.1). Below that, the user selects their competence class(es) based on their expertise. This selection filters all available species to include only those from the selected classes, listed below. In addition, the selected sampling strategy will only consider species from the selected classes. The annotator indicates the presence of a species by checking the corresponding box in the list. Using the 'Search for species' input, the list is filtered by the search string to accelerate annotation. When using the 'validate' sampling strategy, model suggestions are displayed for the species most likely present in the sample, along with their corresponding probabilities. New species can be added and assigned to a sample using the respective input field and 'ADD' button. When using 'validate' or 'refine' as sampling strategy, it is possible to select specific species from a dropdown menu. While for 'validate' the selected species are ignored during the sample selection process, 'refine' will only take into account the selected species, if any, and otherwise use all species from the selected competence classes. The 'SUBMIT' button saves the annotations, the 'SKIP' button discards them.

ACKNOWLEDGMENTS

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

[1] Bradley J. Cardinale, J. Emmett Duffy, Andrew Gonzalez, et al. 2012. Biodiversity loss and its impact on humanity. *Nature* 486 (2012), 59–67. https://api.semanticscholar.org/CorpusID:4333166

[2] Juan Cañas, María Toro-Gómez, Larissa Sugai, et al. 2023. A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring. *Scientific Data* 10, 1 (2023), 771. https://doi.org/10.1038/s41597-023-02666-2

[3] Jurriaan De Vos, Lucas Joppa, John Gittleman, et al. 2015. Estimating the normal background rate of species extinction. *Conservation biology* 29, 2 (2015), 452–462.

[4] Thiago S. Gouvêa, Ilira Troshani, Marc Herrlich, and Daniel Sonntag. 2022. Annotating Sound Events Through Interactive Design of Interpretable Features. In *HHAI 2022: Augmenting Human Intellect - Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence, Amsterdam, The Netherlands, 13-17 June 2022 (Frontiers in Artificial Intelligence and Applications, Vol. 354)*. IOS Press, 305–306. https://doi.org/10.3233/FAIA220225

[5] Thiago Gouvêa, Hannes Kath, Ilira Troshani, et al. 2023. 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*. ijcai.org, Macau, SAR China, 6405–6413. https://doi.org/10.24963/ijcai.2023/711

[6] IPBES. 2019. *Summary for policymakers of the global assessment report on biodiversity and ecosystem services*. Technical Report. Zenodo. https://doi.org/10.5281/ZENODO.3553579 Version Number: summary for policy makers.

[7] Stefan Kahl, Connor Wood, Maximilian Eibl, and Holger Klinck. 2021. BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics* 61 (2021), 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

[8] Hannes Kath, Thiago Gouvêa, and Daniel Sonntag. 2023. A Human-in-the-Loop Tool for Annotating Passive Acoustic Monitoring Datasets. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*. Macau, SAR China, 7140–7144. https://doi.org/10.24963/ijcai.2023/835

[9] Hannes Kath, Thiago S. Gouvêa, and Daniel Sonntag. 2024. Active and Transfer Learning for Efficient Identification of Species in Multi-Label Bioacoustic Datasets. In *Proceedings of the 2024 ACM Conference on Information Technology for Social Good, GoodIT 2024, Bremen, Germany, September 4-6, 2023*. ACM.

[10] Hannes Kath, Patricia P. Serafini, Ivan B. Campos, et al. 2024. Demo: Enhancing Wildlife Acoustic Data Annotation Efficiency through Transfer and Active Learning. In *Proceedings of the 33nd International Joint Conference on Artificial Intelligence, IJCAI*.

[11] Hannes Kath, Patricia P. Serafini, Ivan B. Campos, et al. 2024. Leveraging transfer learning and active learning for data annotation in passive acoustic monitoring of wildlife. *Ecological Informatics* 82 (2024), 102710. https://doi.org/10.1016/j.ecoinf.2024.102710

[12] Bengt Lüers, Patricia P. Serafini, Ivan Braga Campos, and others Gouvêa. 2024. 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.

[13] Félix Michaud, Jérôme Sueur, Maxime LE Cesne, and Sylvain Haupert. 2022. Unsupervised classification to improve the quality of a bird song recording dataset. *ArXiv* abs/2302.07560 (2022). https://api.semanticscholar.org/CorpusID:254625204

[14] Sean Perry, Vaibhav Tiwari, Nishant Balaji, et al. 2021. 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*. IEEE, 633–638. https://doi.org/10.1109/MASS52906.2021.00091

[15] S. Pimm, C. Jenkins, R. Abell, et al. 2014. The biodiversity of species and their rates of extinction, distribution, and protection. *Science* 344, 6187 (2014), 1246752.

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

[17] Alba Solsona-Berga, Kaitlin E. Frasier, Simone Baumann-Pickering, et al. 2020. DetEdit: A graphical user interface for annotating and editing events detected in long-term acoustic monitoring data. *PLoS Comput. Biol.* 16, 1 (2020). https://doi.org/10.1371/journal.pcbi.1007598

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

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

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

[21] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. 2020. Label Studio: Data labeling software. https://github.com/heartexlabs/label-studio

[22] Ilira Troshani, Thiago Gouvea, and Daniel Sonntag. 2024. Leveraging Weakly Supervised and Multiple Instance Learning for Multi-label Classification of Passive Acoustic Monitoring Data. In *KI 2024: Advances in Artificial Intelligence. German Conference on Artificial Intelligence (KI-2024), 47th German Conference on AI, Würzburg, Germany, September 25–27, 2023, located at 47th German Conference on AI, September 25-27, Würzburg, Germany*. Springer.

[23] Devis Tuia, Benjamin Kellenberger, Sara Beery, et al. 2022. Perspectives in machine learning for wildlife conservation. *Nature Communications* 13, 1 (Feb. 2022), 792. https://doi.org/10.1038/s41467-022-27980-y Number: 1, Publisher: Nature Publishing Group.

[24] Juan Sebastián Ulloa, Sylvain Haupert, Juan Felipe Latorre, et al. 2021. scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. *Methods in Ecology and Evolution* (Sept. 2021), 2041–210X.13711. https://doi.org/10.1111/2041-210X.13711

[25] E.O. Wilson and Harvard University (Eds.). 1988. *Biodiversity*. The National Academies Press, Washington, DC.

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