Received: 11 October 2024 | Accepted: 8 November 2024

DOI: 10.1111/2041-210X.14468

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

APPLICATION

Whombat: An open-source audio annotation tool for machine learning assisted bioacoustics

Santiago Martínez Balvanera[1] | Oisin Mac Aodha[2] | Matthew J. Weldy[3,4] | Holly Pringle[1] | Ella Browning[1,5] | Kate E. Jones[1]

[1]Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, UK

[2]School of Informatics, University of Edinburgh, Edinburgh, UK

[3]Department of Forest Ecosystems and Society, College of Forestry, Oregon State University, Corvallis, Oregon, USA

[4]Pacific Northwest Research Station, USDA Forest Service, Corvallis, Oregon, USA

[5]Bat Conservation Trust, Cloisters Business Centre, London, UK

Correspondence
Santiago Martínez Balvanera
Email: santiago.balvanera.20@ucl.ac.uk

Funding information
Consejo Nacional de Humanidades, Ciencias y Tecnologías, Grant/Award Number: 2020-000017-02EXTF-00334

Handling Editor: Camille Desjonquères

Abstract

1. Automated analysis of bioacoustic recordings using machine learning (ML) methods has the potential to greatly scale biodiversity monitoring efforts. The use of ML for high-stakes applications, such as conservation and scientific research, demands a data-centric approach with a focus on selecting and utilizing carefully annotated and curated evaluation and training data that are relevant and representative. Creating annotated bioacoustic datasets presents a number of challenges, such as managing large collections of recordings with associated metadata, developing flexible annotation tools that can accommodate the diverse range of vocalization profiles of different organisms and addressing the scarcity of expert annotators.
2. We present Whombat, a user-friendly, browser-based interface for managing audio recordings and annotation projects, with several visualization, exploration and annotation tools. It enables users to quickly annotate, review, and share annotations, as well as visualize and evaluate a set of machine learning predictions on a dataset. The tool facilitates an iterative workflow where user annotations and machine learning predictions feedback to enhance model performance and annotation quality.
3. We demonstrate the flexibility of Whombat by showcasing two distinct use cases: (1) a project aimed at enhancing automated UK bat call identification at the Bat Conservation Trust (BCT), and (2) a collaborative effort among the USDA Forest Service and Oregon State University researchers exploring bioacoustic applications and extending automated avian classification models in the Pacific Northwest, USA.
4. Whombat is a flexible tool that can effectively address the challenges of annotation for bioacoustic research. It can be used for individual and collaborative work, hosted on a shared server or accessed remotely, or run on a personal computer without the need for coding skills. The code is open-source, and we provide a user guide.

KEYWORDS
AI, audio annotation, bioacoustics, bioinformatics, machine learning, software, sound event detection, visualisation

1 | INTRODUCTION

Recent advancements in Machine Learning (ML) are revolutionizing our ability to analyse large datasets generated by passive acoustic recorders for ecologically relevant signals (Kitzes et al., 2021; Tuia et al., 2022). Open-source Deep Learning models, such as BirdNET (Kahl et al., 2021) and NABat ML (Khalighifar et al., 2022), can be used to monitor birds and bats at scale across large regions. While considerable attention has been directed towards developing sophisticated ML systems, it is crucial to acknowledge the pivotal role of data and the various tasks encompassed within data work in establishing reliable ML implementations (Sambasivan et al., 2021). These tasks include discovery, capture, curation, design and creation of data, which collectively contribute to the quality and effectiveness of machine learning models (Muller et al., 2019). In line with this, the data-centric approach has gained increasing relevance (Jarrahi et al., 2022), emphasizing the collection, curation and management of high-quality training and evaluation data to comprehensively assess model performance and ensure reliability, particularly in high-stakes applications such as conservation. Data work is inherently complex, and audio annotation, encompassing the identification of the location of relevant sound events in audio recordings and the assignment of appropriate labels, represents a time-consuming and labour-intensive process (Cartwright et al., 2019). Often, the creation of ML-ready datasets relies on software tools and technical infrastructure to ease management and enhance efficiency (Reichert et al., 2021; Roe et al., 2021). However, while the broader ML community has recognized the importance of providing accessible, efficient and open-source tools for dataset curation and annotation (Neves & Ševa, 2020; Sager et al., 2021), the bioacoustics community has lagged behind (Stowell, 2022; Tuia et al., 2022).

The annotation process is an integral part of an iterative workflow aimed at continually improving and monitoring the performance of ML models and data quality (Hohman et al., 2020). The evaluation of ML models can help identify errors and areas for potential improvement, such as annotation or data gaps, thereby increasing confidence in the performance of the model (Nahar et al., 2022). Continual annotation of novel data is crucial to monitor the performance of ML models, particularly when exposed to unknown environments, as these can pose a risk to model accuracy and reliability (Saria & Subbaswamy, 2019). However, existing annotation tools often lack appropriate design for effective annotation and machine learning development, hindering the seamless execution of this valuable feedback loop (Table 1).

Creating ML-ready datasets for bioacoustic research is a collaborative effort (Zhang et al., 2020) that requires a combination of modelling, analysis, annotation work, and quality assurance (Jarrahi et al., 2022; Muller et al., 2019). Annotation can be accelerated if tackled by teams working simultaneously and distributing the workload among members with specialized and expert knowledge (Cartwright et al., 2019; Muller et al., 2021). However, managing large collections of audio recordings in bioacoustic research can be overwhelming (Kvsn et al., 2020), as they often contain hundreds or thousands of recordings (Zhang et al., 2013), each with its own set of metadata such as

TABLE 1 Comparison of seven popular software used for acoustic annotation (see Supporting Information for further details).

	Whombat	Arbimon	AvianZ	Kaleidoscope	Label studio	Raven	Sonic visualiser
Open-source	✓	–	✓	–	✓	–	✓
Self-host	✓	–	✓	✓	✓	✓	✓
Collaborative	✓	✓	–	–	✓	–	–
Large datasets	✓	✓		–	✓	–	–
Rich metadata	✓	✓	–	✓	–	–	–
Search capabilities	✓	✓	–	–	–	–	–
Annotation exploration	✓	✓	–	–	–	–	–
Flexible spectrogram	✓	–	✓	✓	–	✓	✓
Flexible annotations	✓	✓	✓	–	–	–	–
Quality assurance	✓	✓	–	–	–	–	–
Training tools	✓	–	–	–	–	–	–
Prediction evaluation	✓	✓	✓	–	–	–	–
Export annotations	✓	–	–	–	✓	✓	✓
Integrated detectors	–	✓	✓	✓	–	–	–

Note: Dashes indicate that the corresponding feature is not supported by the software (to the best of the authors' knowledge), while a checkmark indicates its availability in the iterative workflow of the Whombat annotation tool.

location, date and time of recording, as well as other relevant contextual information. Storing the associated metadata is desired as it can influence modelling decisions and provide contextual cues for acoustic identification (Kshirsagar et al., 2021; Paullada et al., 2021). Being able to locate specific recordings or annotations within these collections is crucial for effective analysis and research but can be time-consuming and difficult without proper tools (Kandel et al., 2012). Providing a platform for collaborative annotation requires finding a balance between accessibility, simplicity and the ability to manage complex and diverse workflows (Simpson et al., 2014).

Bioacoustic annotation is a challenging task due to the wide variety of organisms and vocalization profiles that are studied in bioacoustic research (Odom et al., 2021; Stowell, 2022). Some animals produce long duration and broad-band sounds, while others produce vocalizations that can be clearly localized both in time and frequency. Substantial expertise in the acoustic identification of the target animal is often required and acquiring this knowledge can be a challenging process, often requiring extensive field experience. The pool of bioacoustic experts per taxon is, therefore, typically small and their expert annotation time is valuable (Nahar et al., 2022). While existing annotated data can serve as valuable reference material for training, the process of up-skilling annotators often requires structured guidance and a systematic presentation of diverse target sounds. Existing annotation tools, though possessing many components suitable for training, lack features specifically tailored for this purpose. Additionally, in order to effectively accommodate the varying characteristics of different types of biological sounds, annotation tools must be flexible in terms of their visualization and annotation capabilities (Stowell, 2022). Furthermore, generic audio annotation tools are primarily focused on the analysis of human speech or music and lack the necessary visual representation of audio and consideration of recording context. Conversely, specialized bioacoustic software has often focused on specific taxonomic groups (Marsland et al., 2019; Szewczak, 2010), making it difficult to use these tools for the analysis of other groups. Despite the availability of a variety of annotation tools, none have been able to fully address the complexity of challenges that are inherent to bioacoustic research (Table 1; see Supporting Information for a thorough evaluation of audio annotation tools).

Here we present Whombat, a flexible tool specifically designed to accelerate bioacoustic ML research by facilitating the curation of annotated acoustic datasets. Whombat offers a user-friendly browser-based interface that enables efficient management of acoustic datasets and annotation projects. It provides various visualization, exploration and annotation tools that allow users to annotate, review and share annotations with ease. Moreover, these exploration tools can be employed to visualize, evaluate and explore ML predictions on annotated datasets. Whombat supports an iterative workflow (Figure 1), where user annotations and ML predictions continuously enhance both model performance and annotation quality. Additionally, Whombat is designed to support both individual and collaborative work, enabling hosting on shared servers, cloud platforms or private premises with remote accessibility. Notably, it can also run on personal computers without internet access. The application code is open-source and available at https://github.com/mbsantiago/whombat. To ensure accessibility for all users, we have bundled the tool into executable files for Windows, macOS and Ubuntu, eliminating the need for dependency installation or coding skills. By making Whombat open-source and easily accessible, we aim to empower researchers in bioacoustic ML research and foster advancements in the field.

2 | WHOMBAT FEATURES

In this section we provide a brief description of the features and interface of Whombat, following the order of the intended annotation workflow (Figure 1). This includes the initial setup and loading of data, visualization and navigation tools, annotation capabilities, quality control features and ML model evaluation. Through this overview, we demonstrate how Whombat can enhance the efficiency and accuracy of bioacoustic annotation.

2.1 | Dataset management

The workflow begins by creating an acoustic dataset (Figure 1). A dataset can be created by selecting all recordings within a folder or by importing a pre-existing dataset. The tool supports various audio file formats, including popular lossless formats in Bioacoustics such as WAV and FLAC, as well as the lossy MP3 format and others. Multiple datasets can be managed simultaneously.

Basic media information is scanned and stored for each recording, including its duration, number of channels and sample rate (Figure 2). Whombat also allows the retrieval of metadata from commonly used autonomous recording units, for example Wildife Acoustics and AudioMoth (Hill et al., 2019). Users can edit the location and date-time of recordings on a per-recording basis or import this information from CSV files. Additionally, recordings can be tagged with multiple key-value pairs, providing contextual information relevant to the annotation process. For example, a recording can be tagged with key-value pairs, like species:Myotis lucifugus, sex:Male, age:Adult, and habitat:Forest, to describe the recording target and context. In essence, a key-value pair is a simple way to store data where one piece of information acts as a label (key) and another piece holds the corresponding value. Here, 'species' is the key, and *Myotis lucifugus* is the value associated with that key. This approach allows for flexible, organized and extensible metadata management.

To explore datasets, users can listen to recordings and visualize their spectrograms. Whombat uses spectrograms as the main visualization tool, as they facilitate the quick identification of sound events (Cartwright et al., 2019). Spectrogram parameters and other visual settings are configurable to best suit target sounds. Whombat dynamically generates spectrogram sections on the fly, optimizing computational efficiency and preventing excessive memory usage for long recordings. This allows for easy navigation using scroll bars, eliminating the need to compute and store large spectrograms in

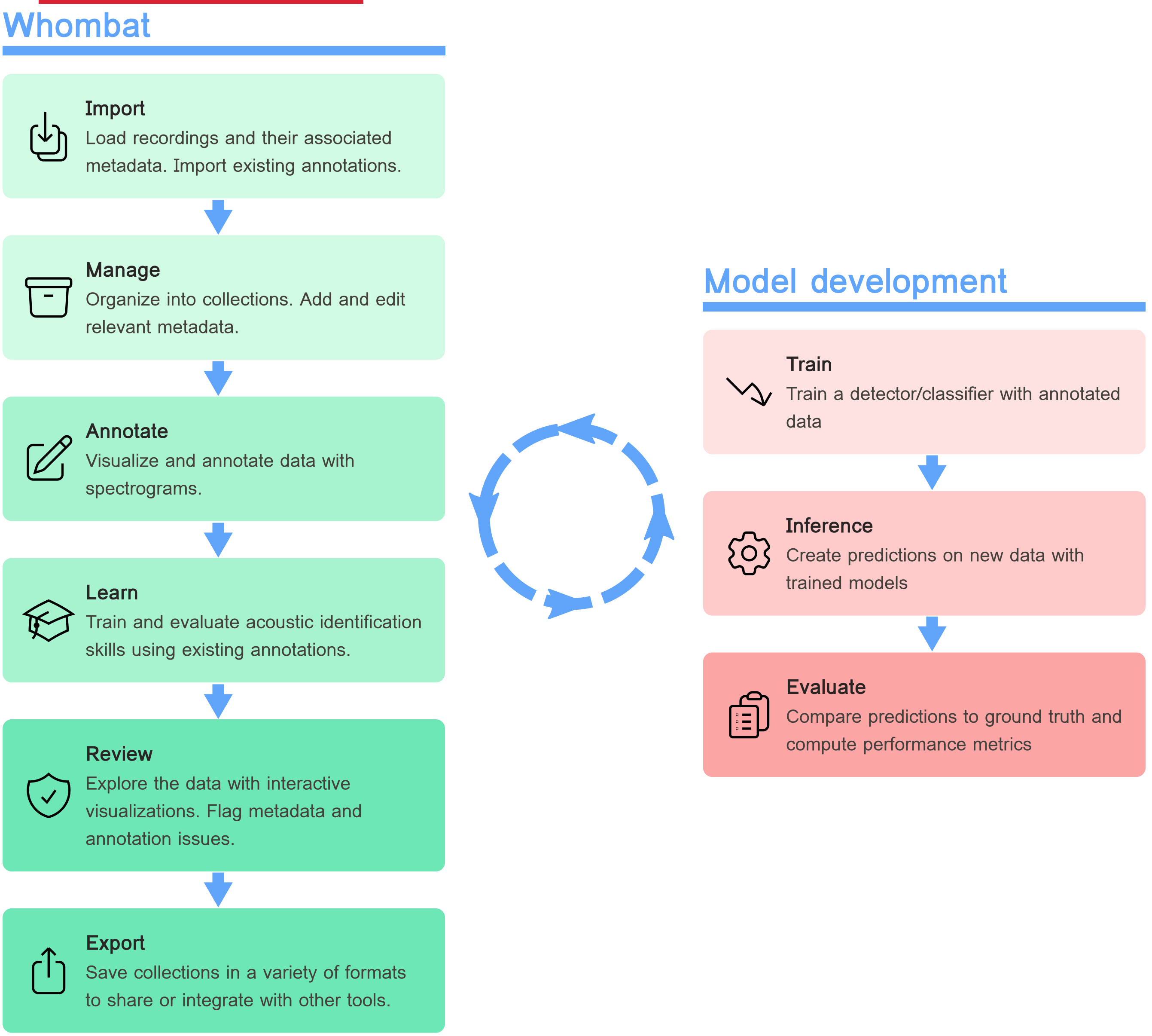

FIGURE 1 The iterative workflow of the Whombat annotation tool. The app enables a feedback loop between user annotations and machine learning predictions, enhancing both model performance and annotation quality. The capabilities of the tool are represented by the green boxes on the left, while the red boxes on the right illustrate the steps in the model development workflow. The arrows indicate the typical direction of the workflow, but the tool provides flexibility for users to navigate between steps. Dashed arrows indicate potential crossover between model and data development. Whombat allows users to export annotations for training machine learning models and then import predictions for comparison with existing annotations. This two-way flow of information empowers users to explore and integrate both components for their analysis.

their entirety. Users can zoom in to relevant parts of the spectrogram or zoom out to scan for interesting sounds. Whombat also provides searching, filtering and sorting tools to quickly browse the recordings of interest.

2.2 | Annotation

Annotation projects can be created by selecting any number of audio clips from recordings of interest. Audio clips are continuous sections extracted from recordings. They can vary in duration and are not constrained to match the length of the original recording. The use of audio clips as the basis of annotation tasks allows cutting the recordings into clips of standardized duration and possibly annotating only a subset of all audio clips. The included clips can be selected within the tool or imported from a CSV file. To create an annotation project, a name, description and annotation instructions for the annotators should be provided.

Once an annotation project is created, each audio clip can then be visualized and annotated. A configurable spectrogram of the clip is displayed, along with recording metadata to provide context to the annotator (Figure 2). Annotation can proceed in different ways

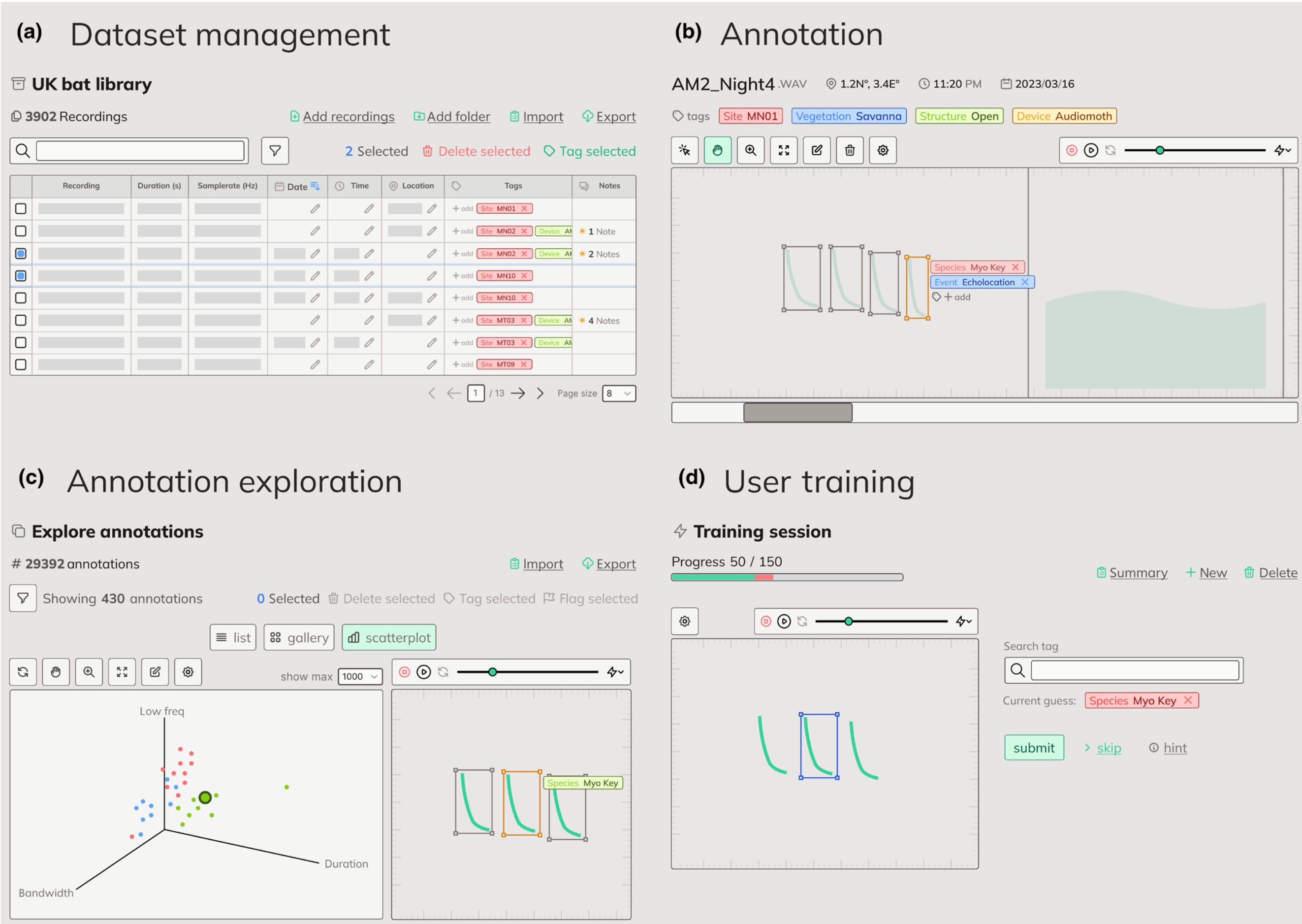

FIGURE 2 Overview of the key features of Whombat. (a) presents dataset management capabilities, including editing, searching and filtering recording metadata. (b) showcases the main annotation interface, offering spectrograms of recordings and various annotation tools. (c) demonstrates annotation exploration features, enabling users to browse existing annotations through filtering and visualizations, such as scatter plots. Finally, (d) highlights the training component of Whombat, where users learn to identify a collection of sound events while receiving guidance on sound event identification.

depending on the project targets and strategy. Users can add any number of key-value tags to the recording clip, for example to specify which species are present within the clip. Relevant sound events can be annotated by locating them within the spectrogram by drawing a vertical line, a temporal interval or a bounding box. Each annotation can be tagged with any amount of key-value pairs, potentially capturing multiple and independent attributes of the sound event, such as species, sound type, sex or the identity of the individual. Although tags can be created freely, we offer a quick search feature to avoid duplication and to ensure consistency.

As annotations progresses, audio clips can be marked as 'ready' once they have been fully annotated according to the project instructions. Annotation progress is tracked by displaying the percentage of audio clips that have been marked as ready, along with the counts of annotated clips and annotations with a given tag. With the aid of filtering and sorting tools, users can focus and prioritize their annotation efforts on specific subsets of the annotation project.

2.3 | Review and exploration

Quality of metadata and annotations can be reviewed and managed through various tools within Whombat. Users can add notes to recordings and annotations to provide additional context and note issues that require fixing. Incomplete annotations can also be flagged, and the issues can be searched to address them efficiently.

In addition, Whombat provides tools for exploring and comparing groups of annotations. The gallery option displays a panel of annotated sound events from different user-selected groups, allowing for easy comparison. For groups of bounding box annotations, the tool can compute statistics on attributes such as the duration, bandwidth and frequency range, and display them in histograms. Whombat also provides an interactive 2D or 3D scatter plot of any combination of the previously mentioned attributes (Figure 2). These visualization tools enable users to become familiar with the variety of sound events and identify potential issues such as outliers and overlaps between categories.

2.4 | User training

Novice users can be incorporated into the workflow by training with existing verified data, which is critical as access to experts in bioacoustics is a recurrent bottleneck for annotation. Our tool addresses this issue by allowing users to learn and improve their annotation skills. The registered annotations can be used to train and evaluate human annotation skills. Users can create training sets by selecting specific annotations, such as those with a particular set of tags. The training sets can be used to conduct training sessions (Figure 2), in which users are shown a series of spectrograms centred at annotated sound events and asked to identify them correctly. After each session, the identification performance is evaluated and displayed, enabling users to track their learning progress and identify areas that need improvement.

2.5 | Data export

Whombat allows users to export their acoustic datasets and annotation projects to multiple file formats. The recommended format is a custom JSON format inspired by the COCO dataset format (Lin et al., 2014), although a CSV format is also available. This makes it possible to use the annotations for training machine learning identification models, other bioacoustic analysis or to share with the wider community.

In addition, the exported dataset and annotation project files can be imported back into the tool. This functionality allows for offline distributed collaborative work where multiple people work on disjointed datasets and share the resulting annotations. This is particularly useful because it bypasses the need for centralized server infrastructure.

2.6 | Closing the loop

To improve ML model performance, the tool provides a way to import model predictions and compare them with user-made annotations. Whombat accepts model predictions in a specific JSON or CSV format, with no restriction on the type of ML model used. Once imported, the set of predictions for a group of recordings is registered as a model run. Users can provide a name and description for the run to help track and organize different model experiments.

Whombat then allows users to evaluate the model run by comparing it with annotations, if available. Several measures of predictive capacity, such as precision and recall, are displayed to help users assess the performance of the model. Users can also explore the predictions using search, sort and filter tools, based on the predicted tag probabilities. This facilitates browsing both success and failure cases, helping to identify potential model improvement opportunities.

In addition to evaluating the ML model, the tool can also be used to diagnose potential data and annotation gaps. By comparing the model predictions with user-made annotations, users can identify cases where the model fails to detect sound events correctly. These cases can then be reviewed to see if there are annotation or data gaps that need to be addressed to improve model performance.

3 | USE CASES

Whombat is designed specifically for audio data in the field of bioacoustics, and its flexibility makes it adaptable to a range of use cases. In this section, we highlight two examples of how the tool can be used: annotation of bat calls from the UK and bird vocalization detection in the Pacific Northwest of the USA. These examples showcase the versatility and potential of the tool for annotating different types of target species and vocalizations.

3.1 | Bat call classification pipeline

The Bat Conservation Trust (BCT) uses Whombat to improve bat detection and classification in the UK. The BCT collects and annotates recordings of bat calls across the UK to enhance the BatDetect tool (Mac Aodha et al., 2018) and advance from bat call detection to a multi-class object detection and classification pipeline (Mac Aodha et al., 2022). Bats play a crucial role in the UK ecosystems (Barlow et al., 2015), and as small, nocturnal, volant mammals that use ultrasonic echolocation for navigation they are routinely monitored using passive acoustic methods (Banner et al., 2018; Barlow et al., 2015; Kerbiriou et al., 2015; Newson et al., 2015; Yoh et al., 2023). Furthermore, interspecific differences in bat echolocation call characteristics enables species or genus level identification from acoustic data. Automating the classification of bat echolocation calls enables monitoring to be carried out at the scales necessary for identifying national conservation management strategies. Improving the detection and classification performance of automated tools, such as BatDetect (Mac Aodha et al., 2022), is therefore crucial for the success of conservation efforts.

Whombat has enabled the BCT to generate precise bat call annotations while offering flexibility in the types of annotations captured. Bat calls are short and high-frequency, making them well-suited for annotation with bounding boxes tightly placed around the main harmonic. Annotators use tags in the form species: *<species>* to indicate the bat species, and event: *<call type>* to specify the call type (e.g. echolocation, social call, feeding buzz). In cases of uncertainty, a generic tag like order: *Chiroptera* can be employed. Additionally, potential false positives can be annotated with an event: *Noise* tag to reduce confusion. While bats are the main focus at the BCT, and it is important to be able to capture their different types of calls, it is also crucial to identify confounding noises and register co-occurring sounds that can be important for downstream analysis.

The tool has enabled the BCT to centralize annotation work, eliminating the complexities of harmonizing previously independent efforts. Furthermore, Whombat has allowed to streamline the review process by allowing to assess the annotator's work. This collaborative approach has proven valuable, leading to the identification and correction of mislabelled annotations due to confusion in the annotation instructions. This has allowed the BCT to improve both the quality and quantity of their annotations. The collaborative nature of the tool also allows for efficient data sharing and analysis, making it an essential tool for BCT and their bat conservation work.

A total of 29 independent datasets of bat recordings, comprising over 70,000 annotated calls, have been processed at the BCT using Whombat. It has been used by more than 15 independent annotators from the BCT and partner institutions. The annotations generated using Whombat directly inform the training of machine learning algorithms (Mac Aodha et al., 2022), demonstrating improved performance compared to other existing bat detection tools. These annotations and the refined models they enable extend the BCT's capability to understand bat population responses to anthropogenic environmental change and inform conservation efforts.

3.2 | Bird song annotation

In 1994, the Northwest Forest Plan was introduced in the United States Pacific Northwest to shift federal land management policies from prioritizing timber harvesting to a more holistic approach that includes protecting and restoring the habitat of old-forest species and biodiversity (Espy & Babbitt, 1994). One of the components of this plan is the long-term monitoring of federally threatened northern spotted owl (*Strix occidentalis caurina*) populations through a two-phase approach (Lint, 1999). The first phase involved estimating vital rates and demographic performance using mark-resight methods on historical territories (Franklin et al., 2021). The second phase began in 2020 and focused on estimating occupancy and habitat models through passive acoustic monitoring (Lesmeister & Jenkins, 2022).

The transition to phase two monitoring is a crucial moment in conserving and managing forested lands in the Pacific Northwest. Not only are spotted owl conservation and management objectives being met (Lesmeister & Jenkins, 2022; Weldy et al., 2023), but the multispecies acoustic monitoring data can also be used to address other conservation, research, or management objectives. To this end, researchers from Oregon State University and the USDA Forest Service are using Whombat to annotate avian sounds (>30,000 annotations) and validate model predictions for various wildlife monitoring programs targeting federally threatened species like the northern spotted owl and the marbled murrelet (*Brachyramphus marmoratus*), as well as sensitive species such as the white-headed woodpecker (*Dryobates albolarvatus*), and supporting broader biodiversity monitoring efforts (>80 species).

The dynamic acoustic and spectrogram adjustments provided by the tool have improved the quality of target species annotation, increased efficiency in reviewing model predictions, and aided in tracking acoustic review and labelling efforts. In addition, the annotation formatting of Whombat is flexible and dynamic, allowing annotators to pursue multiple annotation objectives simultaneously. They can opportunistically collect biophonic examples for non-target species and create hierarchical label structures where sound types are nested within broader categories. These hierarchical labels cascade across increasingly fine-scale taxonomic determinations. Additionally, annotators can label acoustic metadata such as approximate distances or overlapping sound types, which serves to improve model training and enhance the understanding of model performance. The adoption of passive acoustic monitoring represents an important step forward in conserving and managing forested lands in the Pacific Northwest. Using innovative tools such as Whombat enhances these efforts.

4 | DISCUSSION

The modularity and extensibility of Whombat enables many opportunities for future development (see the Supporting Information for more details on the software design). We invite the community to contribute to its growth and suggest potential areas of improvement, such as the ability to group annotations into sequences, model comparison and data iteration visualizations, and dashboards for ecological insights and quick exploration. One possibility for expanding the user base of the tool is to incorporate a citizen science approach by evaluating the reliability of user annotations. We believe these and other potential directions will help make Whombat an even more powerful tool for bioacoustic research and conservation efforts.

Unlike other annotation solutions (e.g. Marsland et al., 2019), our tool does not include embedded machine learning detectors. We made this decision to simplify the software and decouple the annotation process from the development and maintenance of machine learning models. Instead, our focus is on providing a user-friendly interface for efficient and accurate annotation. We also provide an interface for importing and exporting model predictions, allowing users to incorporate their own machine learning models into their annotation projects. Additionally, the tool allows to export annotations in a format that is compatible with training frameworks for bioacoustic detection models (e.g. Mac Aodha et al., 2022).

By providing an accessible, open-source tool for bioacoustic annotation, we hope to empower research teams to generate high-quality acoustic datasets for their projects, including those without extensive coding experience. The modular and extensible design of the software allows for customization to meet individual project needs and encourages community involvement in the development of new features. By lowering the barrier to entry for annotation projects, we aim to foster the creation of diverse and shareable datasets that can advance research in bioacoustics.

AUTHOR CONTRIBUTIONS

Santiago Martínez Balvanera, Oisin Mac Aodha and Kate E. Jones conceived the project and designed the software. Santiago Martínez Balvanera wrote the software, the documentation and the first draft of the manuscript. Ella Browning, Holly Pringle and Matthew J. Weldy led annotation efforts described in the use cases and provided extensive feedback on the software usability. Kate Jones and Oisin Mac Aodha reviewed the manuscript and provided feedback on the software design. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

We thank Giada Giacometti and other team members from the BCT, Sospeter Kiwibot and Paul Webala for beta testing and valuable feedback. We also thank Grant Van Horn, Everardo Gustavo Robredo Ezquivelzeta, Veronica Guiterrez-Zamora and Cristina MacSwiney for helpful discussions. Financial support was provided by Consejo Nacional de Ciencia y Tecnologia (CONACYT) Award Number 2020-000017-02EXTF-00334.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210X.14468.

DATA AVAILABILITY STATEMENT

Whombat can be downloaded for Windows, macOS and Linux from https://github.com/mbsantiago/whombat/releases. The source code is available at https://doi.org/10.5281/zenodo.10604169 (Martínez Balvanera, 2024) or at https://github.com/mbsantiago/whombat, and the README provides installation instructions. The Software guide, which provides detailed usage instructions, is available in English at https://mbsantiago.github.io/whombat/. Sample datasets mentioned in the article are included in the GitHub repository.

ORCID

Santiago Martínez Balvanera https://orcid.org/0000-0003-1214-1938
Oisin Mac Aodha https://orcid.org/0000-0002-5787-5073
Matthew J. Weldy https://orcid.org/0000-0002-8627-2466
Ella Browning https://orcid.org/0000-0002-7959-9292
Kate E. Jones https://orcid.org/0000-0001-5231-3293

REFERENCES

Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M., & Litt, A. R. (2018). Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecology and Evolution*, *8*(12), 6144–6156. https://doi.org/10.1002/ece3.4162

Barlow, K. E., Briggs, P. A., Haysom, K. A., Hutson, A. M., Lechiara, N. L., Racey, P. A., Walsh, A. L., & Langton, S. D. (2015). Citizen science reveals trends in bat populations: The National bat Monitoring Programme in Great Britain. *Biological Conservation*, *182*(February), 14–26. https://doi.org/10.1016/j.biocon.2014.11.022

Cartwright, M., Dove, G., Mendez, A. E. M., Bello, J. P., & Nov, O. (2019). Crowdsourcing multi-label audio annotation tasks with citizen scientists. In *International conference on human factors in computing systems, May, 292*. Association for Computing Machinery. https://doi.org/10.1145/3290605.3300522

Espy, M., & Babbitt, B. (1994). Record of decision for amendments to forest service and bureau of land management planning documents within the range of the northern spotted owl. In *US Department of Agriculture, Forest Service and US Department of the Interior, Bureau of Land Management*. United States Department of Agriculture Natural Resources Conservation Service.

Franklin, A. B., Dugger, K. M., Lesmeister, D. B., Davis, R. J., David Wiens, J., White, G. C., Nichols, J. D., Hines, J. E., Yackulic, C. B., Schwarz, C. J., Ackers, S. H., Andrews, L. S., Bailey, L. L., Bown, R., Burgher, J., Burnham, K. P., Carlson, P. C., Chestnut, T., Conner, M. M., … Wise, H. (2021). Range-wide declines of northern spotted owl populations in the Pacific northwest: A meta-analysis. *Biological Conservation*, *259*(July), 109168. https://doi.org/10.1016/j.biocon.2021.109168

Hill, A. P., Prince, P., Snaddon, J. L., Patrick Doncaster, C., & Rogers, A. (2019). AudioMoth: A low-cost acoustic device for monitoring biodiversity and the environment. *HardwareX*, *6*(October), e00073. https://doi.org/10.1016/j.ohx.2019.e00073

Hohman, F., Wongsuphasawat, K., Kery, M. B., Patel, K., & Patel, K. (2020). Understanding and visualizing data iteration in machine learning. In *International conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376177

Jarrahi, M. H., Memariani, A., & Guha, S. (2022). The principles of data-centric AI (DCAI). arXiv https://doi.org/10.48550/arxiv.2211.14611

Kahl, S., Wood, C. M., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, *61*, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236

Kandel, S., Paepcke, A., Hellerstein, J. M., & Heer, J. (2012). Enterprise data analysis and visualization: An interview study. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2917–2926. https://doi.org/10.1109/tvcg.2012.219

Kerbiriou, C., Julien, J. F., Bas, Y., Marmet, J., Lorrilliere, R., Azam, C., Gasc, A., & Lois, G. (2015). Vigie-Chiro: 9 ans de suivi des tendances des espèces communes. *Symbioses*, *34*, 35.

Khalighifar, A., Gotthold, B. S., Adams, E., Barnett, J., Beard, L. O., Britzke, E. R., Burger, P. A., Chase, K., Cordes, Z., Cryan, P. M., Ferrall, E., Fill, C. T., Gibson, S. E., Haulton, G. S., Irvine, K. M., Katz, L. S., Kendall, W. L., Long, C. A., Mac Aodha, O., … Reichert, B. E. (2022). NABat ML: Utilizing deep learning to enable crowdsourced development of automated, scalable solutions for documenting north American bat populations. *Journal of Applied Ecology*, *59*, 2849–2862. https://doi.org/10.1111/1365-2664.14280

Kitzes, J., Blake, R. E., Bombaci, S. P., Chapman, M., Durán, S. M., Huang, T., Joseph, M. B., Lapp, S., Marconi, S., Oestreich, W. K., Rhinehart, T. A., Schweiger, A. K., Song, Y., Surasinghe, T., & Di Yang, K. Y. (2021). Expanding NEON biodiversity surveys with new instrumentation and machine learning approaches. *Ecosphere*, *12*(11), e03795. https://doi.org/10.1002/ecs2.3795

Kshirsagar, M., Robinson, C., Yang, S., Gholami, S., Klyuzhin, I. S., Mukherjee, S., Nasir, M., Ortiz, A., Oviedo, F., Tanner, D., Trivedi, A., Xu, Y., Zhong, M., Dilkina, B., Dodhia, R., & Lavista Ferres, J. M. (2021). "Becoming good at AI for good." *AAAI/ACM Conference on AI, Ethics, and Society*, 664–673. https://doi.org/10.1145/3461702.3462599

Kvsn, R. R., Montgomery, J., Garg, S., Charleston, M. A., & Charleston, M. (2020). Bioacoustics data analysis a taxonomy, survey and open challenges. *IEEE Access*, *8*(March), 57684–57708. https://doi.org/10.1109/access.2020.2978547

Lesmeister, D. B., & Jenkins, J. M. A. (2022). Integrating new technologies to broaden the scope of northern spotted owl monitoring and linkage with USDA Forest Inventory Data. *Frontiers in Forests and Global Change*, *5*, 966978. https://doi.org/10.3389/ffgc.2022.966978

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Lawrence Zitnick, C. (2014). Microsoft Coco: Common objects in context. In *Computer Vision 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, part V 13* (pp. 740–755). Springer.

Lint, J. (1999). *Northern spotted owl effectiveness monitoring plan for the northwest Forest plan* (Vol. *440*). US Department of Agriculture, Forest Service, Pacific Northwest Research Station.

Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., Harder, B., Kinsey, L., Mead, G. R., Newson, S. E., Pandourski, I., Parsons, S., Russ, J., Szodoray-Paradi, A., Szodoray-Paradi, F., Tilova, E., Girolami, M., Brostow, G., & Jones, K. E. (2018). Bat detective learning tools for bat acoustic signal detection. *PLoS Computational Biology*, *14*(3), e1005995. https://doi.org/10.1371/journal.pcbi.1005995

Mac Aodha, O., Martínez Balvanera, S., Damstra, E., Cooke, M., Eichinski, P., Browning, E., Barataud, M., Boughey, K., Coles, R., Giacomini, G., Cristina Mac Swiney G, M., Obrist, M. K., Parsons, S., Sattler, T., & Jones, K. E. (2022). Towards a general approach for bat echolocation detection and classification. *bioRxiv* https://doi.org/10.1101/2022.12.14.520490

Marsland, S., Marsland, S., Priyadarshani, N., Juodakis, J., Juodakis, J., & Castro, I. (2019). AviaNZ: A future-proofed program for annotation and recognition of animal sounds in Long-time field recordings. *Methods in Ecology and Evolution*, *10*(8), 1189–1195. https://doi.org/10.1111/2041-210x.13213

Martínez Balvanera, S. (2024). mbsantiago/whombat: v0.7.0 (v0.7.0). *Zenodo*. https://doi.org/10.5281/zenodo.13910236

Muller, M., Lange, I., Wang, D., Piorkowski, D., Jason, T., Vera Liao, Q., Dugan, C., & Erickson, T. (2019). How data science workers work with data: Discovery, capture, curation, design, creation. In *International conference on human factors in computing systems* (p. 126). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300356

Muller, M., Wolf, C. T., Andres, J., Desmond, M., Joshi, N. N., Ashktorab, Z., Sharma, A., Brimijoin, K., Pan, Q., Duesterwald, E., & Dugan, C. (2021). Designing ground truth and the social life of labels. In *International Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3411764.3445402

Nahar, N., Zhou, S., Lewis, G., & Kästner, C. (2022). Collaboration challenges in building ml-enabled systems: Communication, documentation, engineering, and process. In *Proceedings of the 44th international conference on software engineering* (pp. 413–425). Association for Computing Machinery. https://doi.org/10.1145/3510003.3510209

Neves, M., & Ševa, J. (2020). Annotationsaurus: A searchable directory of annotation tools. *arXiv* https://doi.org/10.48550/arXiv.2010.06251

Newson, S. E., Evans, H. E., & Gillings, S. (2015). A novel citizen science approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England. *Biological Conservation*, *191*(November), 38–49. https://doi.org/10.1016/j.biocon.2015.06.009

Odom, K. J., Araya-Salas, M., Morano, J. L., Ligon, R. A., Leighton, G. M., Taff, C. C., Dalziell, A. H., Billings, A. C., Germain, R. R., Pardo, M., de Andrade, L. G., Hedwig, D., Keen, S. C., Shiu, Y., Charif, R. A., Webster, M. S., & Rice, A. N. (2021). Comparative bioacoustics: A roadmap for quantifying and comparing animal sounds across diverse taxa. *Biological Reviews of the Cambridge Philosophical Society*, *96*(4), 1135–1159. https://doi.org/10.1111/brv.12695

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, *2*(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Reichert, B. E., Bayless, M. L., Cheng, T. L., Cheng, T. L., Coleman, J. T. H., Francis, C. M., Frick, W. F., Gotthold, B. S., Irvine, K. M., Lausen, C., Li, H., Loeb, S. C., Reichard, J. D., Rodhouse, T. J., Segers, J. L., Siemers, J. L., Thogmartin, W. E., & Weller, T. J. (2021). NABat: A top-down, bottom-up solution to collaborative continental-scale monitoring. *Ambio: A Journal of the Human Environment*, *50*(4), 901–913. https://doi.org/10.1007/s13280-020-01411-y

Roe, P., Eichinski, P., Fuller, R. A., McDonald, P. G., Schwarzkopf, L., Towsey, M., Truskinger, A., Tucker, D., & Watson, D. M. (2021). The Australian acoustic observatory. *Methods in Ecology and Evolution*, *12*(10), 1802–1808. https://doi.org/10.1111/2041-210X.13660

Sager, C., Janiesch, C., & Zschech, P. (2021). A survey of image labelling for computer vision applications. *Journal of Business Analytics*, *4*(2), 91–110. https://doi.org/10.1080/2573234X.2021.1908861

Sambasivan, N., Shivani Kapania, H., Highfill, D. A., Paritosh, P. K., & Aroyo, L. (2021). 'Everyone wants to do the model work, not the data work': Data cascades in high-stakes AI. *International Conference on Human Factors in Computing Systems*. https://doi.org/10.1145/3411764.3445518

Saria, S., & Subbaswamy, A. (2019). Tutorial: Safe and reliable machine learning. arXiv: Learning.

Simpson, R. J., Simpson, R. M., Page, K. R., & De Roure, D. (2014). Zooniverse: Observing the World's largest citizen science platform. In *The Web Conference* (pp. 1049–1054). Association for Computing Machinery. https://doi.org/10.1145/2567948.2579215

Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*, *10*, e13152. https://doi.org/10.7717/peerj.13152

Szewczak, J. M. (2010). SonoBat v.3. SonoBat. https://sonobat.com/

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, *13*(1), 792. https://doi.org/10.1038/s41467-022-27980-y

Weldy, M. J., Lesmeister, D. B., Yackulic, C. B., Appel, C. L., McCafferty, C., & David Wiens, J. (2023). Long-term monitoring in transition: Resolving spatial mismatch and integrating multistate occupancy data. *Ecological Indicators*, *146*, 109815. https://doi.org/10.1016/j.ecolind.2022.109815

Yoh, N., Seaman, D. J. I., Deere, N. J., Bernard, H., Bicknell, J. E., & Struebig, M. J. (2023). Benign effects of logging on aerial insectivorous bats in Southeast Asia revealed by remote sensing technologies. *Journal of Applied Ecology*, *60*, 1210–1222. https://doi.org/10.1111/1365-2664.14398

Zhang, A. X., Muller, M., Muller, M., Muller, M., & Wang, D. (2020). How do data science workers collaborate? Roles, workflows, and tools. *Proceedings of the ACM on Human-Computer Interaction*, *4*(May), 1–23. https://doi.org/10.1145/3392826

Zhang, J., Huang, K., Cottman-Fields, M., Truskinger, A., Roe, P., Duan, S., Dong, X., Towsey, M., & Wimmer, J. (2013). *Managing and analysing big audio data for environmental monitoring*. Science & Engineering Faculty. https://doi.org/10.1109/cse.2013.146

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix A. Evaluation of existing audio annotation tools.

Appendix B. Description of software design and principles.

Table SI.1. Percentage of existing annotation tools that satisfy design criteria.

How to cite this article: Martínez Balvanera, S., Mac Aodha, O., Weldy, M. J., Pringle, H., Browning, E., & Jones, K. E. (2024). Whombat: An open-source audio annotation tool for machine learning assisted bioacoustics. *Methods in Ecology and Evolution*, *00*, 1–10. https://doi.org/10.1111/2041-210X.14468