Ecological Informatics 87 (2025) 103026

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

LEAVES: An open-source web-based tool for the scalable annotation and visualisation of large-scale ecoacoustic datasets using cluster analysis

Thomas Napier [a],*, Euijoon Ahn [b], Slade Allen-Ankins [a], Lin Schwarzkopf [a], Ickjai Lee [b]

[a] College of Science and Engineering, James Cook University, Townsville, 4811, Queensland, Australia
[b] College of Science and Engineering, James Cook University, Cairns, 4870, Queensland, Australia

ARTICLE INFO

Keywords:
Acoustic monitoring
Clustering
Computer software
Decision support
Ecoacoustic analysis
Machine learning application

ABSTRACT

Ecoacoustics has emerged as a pivotal discipline in the conservation and monitoring of ecosystems, offering insights into species' behaviour and ecosystem health through soundscape analysis. Central to this is the need for accurate annotations of environmental audio recordings, which underpin the computational models used in ecological monitoring. However, due to the increasingly large scale of datasets, annotation using existing tools and techniques cannot be performed at feasible speeds or with the necessary accuracy required for real-world application. The LEAVES (Large-scale Ecoacoustics Annotation and Visualisation with Efficient Segmentation) platform addresses this gap by leveraging unsupervised clustering techniques optimised for the high-throughput annotation of large-scale ecoacoustics datasets. Our evaluation across six real-world datasets shows that LEAVES improves annotation efficiency by up to 7.12 times compared to manual annotation while maintaining 79%–90% label similarity to validated data. We expect that our proposed tool will greatly accelerate the annotation process when generating high-quality labelled datasets, supporting larger-scale studies with broader community engagement in ecoacoustics research.

1. Introduction

Globally, the urgent need to monitor and preserve native species grows stronger against the backdrop of deteriorating ecosystems and escalating rates of species loss due to habitat destruction (Chase et al., 2020; Powers and Jetz, 2019; Gonçalves-Souza et al., 2020). By capturing the acoustic signatures of diverse ecosystems, ecoacoustics offers non-invasive insights into biodiversity dynamics and environmental changes (Gibb et al., 2018; Ross et al., 2023; Farina and Gage, 2017). Through the analyses of sounds produced by soniferous animals and the ambient acoustic environment, ecoacoustics can provide insights into species' presence, behaviour, and changes in habitat quality, offering a comprehensive approach to understanding and preserving biodiversity.

Passive Acoustic Monitoring (PAM) complements ecoacoustics by offering a scalable, non-invasive approach to long-term environmental monitoring, reducing the need for labour-intensive field surveys and enabling continuous data collection (Hoefer et al., 2023). It involves strategically deploying low-cost acoustic sensors to passively record vocal fauna, human activity, and natural phenomena. Unlike traditional, laborious manual surveys historically conducted by zoologists and ecologists on site, PAM can be automated and scaled (Gibb et al., 2018; Roe et al., 2021). Although PAM is a powerful framework for studying terrestrial communities and their habitats, it has limitations. In particular, the scale of the generated data poses analytical challenges, as the volume of recordings often exceeds the capacity for human analysis, creating a need for efficient, automated tools (Joshi et al., 2017).

With the advent of Big Data and Machine Learning (ML) technologies, including Deep Learning (DL), significant progress has been made towards their application in ecoacoustics (Stowell, 2022; Dufourq et al., 2022; Quinn et al., 2022; Hao et al., 2022; Guerrero et al., 2023). Integrating ML and DL into ecoacoustics represents a promising frontier for automating and enhancing the analysis of vast acoustic datasets generated by PAM systems. However, the transition to these advanced computational techniques remains hindered by the scarcity of adequately labelled data, which is essential for training the sophisticated large-scale models that underpin DL technologies (Mcloughlin et al., 2019; LeBien et al., 2020; Napier et al., 2024b).

* Corresponding author.
E-mail address: thomas.napier@jcu.edu.au (T. Napier).

https://doi.org/10.1016/j.ecoinf.2025.103026
Received 11 August 2024; Received in revised form 14 January 2025; Accepted 15 January 2025
Available online 19 February 2025

Large-scale ecoacoustics datasets generated from networks like the Australian Acoustic Observatory (A2O) (Roe et al., 2021) and the U.S. Northeast Passive Acoustic Sensing Network (NEPAN) (Van Parijs et al., 2015) pose several unique challenges due to their vast scales, the ongoing generation of continuous recordings from the natural environment, and the inherent complexities of natural soundscapes (Sueur and Farina, 2015). Typically, such datasets feature multi-class, multi-species vocalisations, with high inter-taxa diversity across large communities or biomes (Pijanowski et al., 2011). Exploring these datasets to understand the multiple layers of acoustic information — such as biophony, geophony, and anthrophony — requires specialised functionality that many existing software solutions do not accommodate (Napier et al., 2024a).

Although automated annotation methods, such as those employed by advanced ML applications like BirdNET (Kahl et al., 2021), have made significant strides in processing these datasets, they often necessitate post-processing to ensure the accuracy of the results (Wood and Kahl, 2024). This highlights the indispensable role of human expertise in the refinement of automated annotations, ensuring high accuracy and reliability (Kholghi et al., 2018; Ghani et al., 2023). Given these challenges, this research introduces the LEAVES (**L**arge-scale **E**coacoustics **A**nnotation and **V**isualisation with **E**fficient **S**egmentation) platform (hereafter "the tool") as an efficient and accessible web-based tool designed to accelerate the annotation of large-scale natural soundscape datasets. This tool capitalises on the latest advances in ML and signal processing to create an efficient and scalable solution for ecoacoustic data annotation, incorporating human oversight for improved accuracy. It operates as a decision support system ensuring that the final annotations are validated, thus embodying a human-in-the-loop approach (Mosqueira-Rey et al., 2023).

We aim to provide functionality suitable for users searching for a versatile and efficient platform for audio annotation and visualisation for large dataset pre-filtering and annotation. By focusing on the issues outlined above, the developed system aims to contribute to the burgeoning field of ecoacoustics, providing an innovative solution for use in the annotation and modelling of ecoacoustics. The key innovations of this workflow include:

- Ecoacoustics processing using a tailored approach for annotating large-scale soundscapes efficiently and cost-effectively, ensuring the production of quality labelled sound datasets, including outputs from intermediary processes if required;
- A modular efficiency-optimised labelling system that employs unsupervised segmentation techniques for high-throughput processing, significantly enhancing data annotation speeds;
- Support for various audio formats and a collection of customisation and hyper-parameter tuning options, ensuring adaptability to a range of ecoacoustic research needs;
- 3D embedding visualisations with data-point interaction and real-time spectrogram and waveform analysis, allowing for in-depth exploration of complex natural soundscapes;
- Formulation of a new composite score as an internal clustering validity measure to balance cluster tightness and separation, cluster dispersion, and cluster compactness.

2. Related works

2.1. Requirements of ecoacoustics annotation workflows

Ecoacoustics research demands robust and adaptable workflows capable of processing large and varied datasets. Such workflows include key components like data collection, handling, pre-processing, visualisation, and annotation, each essential for generating usable ecoacoustic datasets (Darras et al., 2023; Stowell, 2022), as depicted in Fig. 1. Large-scale PAM forms the foundation of ecoacoustic studies, involving the collection of extensive raw audio data from diverse ecological settings. However, the lack of standardised protocols and the variability in sensor configurations can negatively impact the consistency and scalability of data collection efforts. Effective unsupervised workflows must therefore incorporate flexible approaches capable of handling diverse data sources and configurations to maintain data quality and comparability across studies.

To this end, efficient data handling is essential for managing PAM datasets, including storage, archiving, and interoperability to meet FAIR (Findable, Accessible, Interoperable, and Reusable) principles, which support data sharing and collaboration in ecoacoustic research (Wilkinson et al., 2016). However, current data handling systems often fail to fully comply with these principles, which limits data accessibility and usability (Villanueva-Rivera and Pijanowski, 2012). Establishing robust data handling mechanisms that align with FAIR principles is essential for supporting large-scale ecoacoustic research and promoting open science (Vella et al., 2022).

Pre-processing is a critical step that transforms raw ecoacoustic data into formats suitable for detailed analysis. This process can include denoising to enhance signal clarity, signal transformation to extract meaningful information, and feature extraction to identify salient characteristics of the acoustic signals (Stowell, 2022). Existing software often lacks the adaptability to efficiently pre-process diverse and complex ecoacoustic signals, which can include overlapping biophony (biological sounds), geophony (non-biological natural sounds), and anthrophony (human-made sounds). Effective pre-processing should be flexible enough to handle these complexities and improve the overall quality of the data used for further analysis.

Visualisation is indispensable for interpreting complex ecoacoustic information. Signal representations such as spectrograms and waveforms are widely used, aiding both experts and non-experts in identifying patterns and anomalies (Towsey et al., 2014). However, many of these existing representations are static and lack interactivity, which limits their effectiveness in engaging users and facilitating deeper analysis. Recent work has highlighted the need for more dynamic and interactive visualisation approaches that can make ecoacoustic data interpretation more intuitive and accessible, particularly for citizen scientists who may not have specialised training (Darras et al., 2023; Napier et al., 2024b).

Annotation is a fundamental aspect of ecoacoustic workflows, involving the labelling of data to create meaningful and reliable datasets for analysis. It can be conducted manually, through citizen science contributions, or with software-assisted approaches. Manual annotation, while accurate, is time-consuming and resource-intensive, whereas citizen science-based approaches can introduce variability in data quality due to differences in participant expertise (Cartwright et al., 2017). Software-assisted methods aim to reduce the manual workload by automating parts of the annotation process; however, many current systems still require significant manual intervention and lack the advanced automation capabilities needed for efficient, large-scale analysis (Stowell, 2022). An effective annotation system must balance automation with expert validation, ensuring that data labelling is both efficient and accurate, thereby supporting the scalability of ecoacoustic research.

2.2. Existing software

Experts traditionally label sounds manually, but the growing ecoacoustic data volume makes manual annotation increasingly untenable (Gan et al., 2021; Stowell and Sueur, 2020). As the field has evolved, there has been a clear pivot in the ecoacoustic community towards leveraging software designed for visualising, annotating, and analysing natural soundscape data. Available software for audio analysis generally falls into two broad categories: general audio editing software and bioacoustics- / ecoacoustics-specific software. Applications like Audacity (Busleiman et al., 2020) offer basic visualisation with spectrograms and waveforms but lack advanced features like

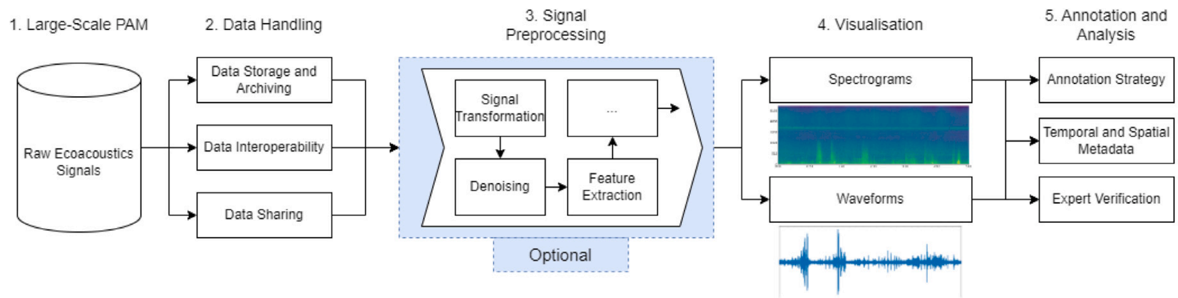


Fig. 1. Overview of the general components in a large-scale ecoacoustics annotation workflow.

Table 1
Comparison of audio annotation and visualisation software based on features relevant to ecoacoustic workflows. Ratings: ○(none or poor), ✧(basic), ★(high quality). Usability levels: Easy (E) for beginner-friendly tools, Moderate (M) for users with basic software knowledge, and Advanced (A) for users with specialised technical expertise.

Feature	Free Tools							Paid Tools					
	LEAVES	Audacity	Sonic Visualiser	Praat	WaveSurfer	Arbimon	ecoSound-web	Sonobat	BatExplorer	BatSound	SASLab Pro	Raven Pro	Kaleidoscope Pro
Data handling	★	★	✧	★	✧	★	★	✧	✧	✧	★	★	★
User interface	★	★	★	✧	✧	★	★	✧	✧	✧	★	★	★
Audio playback	★	★	✧	★	✧	✧	★	✧	✧	✧	★	★	★
Status updates	★	○	✧	✧	○	✧	✧	★	★	★	★	✧	★
Multi-class annotation	★	○	✧	✧	✧	★	★	○	○	○	✧	★	★
High-dimensional visualisation	★	○	○	○	○	○	○	○	○	○	★	○	★
Advanced segmentation	★	○	○	✧	✧	✧	○	✧	✧	✧	✧	✧	★
User defined algorithms	★	○	○	✧	○	✧	○	○	○	○	○	○	✧
Flexible environment support	★	✧	✧	★	✧	✧	✧	○	○	✧	○	✧	★
Efficiency optimised labelling	★	○	○	○	○	✧	○	○	○	○	✧	✧	✧
Usability	M	E	M	M	M	E	E	A	A	A	A	M	A
Reference	This Study	Busleiman et al. (2020)	Cannam et al. (2010)	Boersma and Weenink (2001)	Sjölander and Beskow (2000)	Aide et al. (2013)	Darras et al. (2023)	Szewczak (2010)	BatExplorer (2020)	Pettersson (2004)	Specht (2002)	Charif et al. (2010)	WildlifeAcoustics (2020)

unsupervised segmentation or supervised classification. As a result, while general audio editors are user-friendly and effective for certain tasks, they fall short in handling some of the complexities of large-scale ecoacoustic research, highlighting the need for more specialised solutions.

In contrast, ecoacoustics-specific software is designed to meet these specialised needs, offering features like frequency-time representation, multi-class labelling, and noise reduction. Recent advancements in machine learning have further expanded the potential for automated analysis. For instance, Wang et al. (2023) demonstrated a VGGish-based method for unsupervised classification of biological sound components and their spatio-temporal variations in subtropical forests, highlighting the utility of such approaches for detecting and categorising sound types in complex soundscapes. Despite these advancements and the wide range of general audio software available, relatively few options are specifically designed for large-scale ecoacoustic research. However, one of the stand-out tools in this domain is Kaleidoscope Pro (WildlifeAcoustics, 2020), which is known for its comprehensive analysis capabilities. It is particularly valued for its ability to perform detailed acoustic analyses, including some specific bat and bird species identification through cluster analysis (Nocera et al., 2019; Marchal et al., 2022). Despite these strengths, however, the software's cost can be prohibitive for smaller research groups or individual researchers, limiting its accessibility. Furthermore, its clustering process requires users to deeply understand the specific signal characteristics of the species under study, such as frequency bands, amplitude, call lengths, and more nuanced parameters involved in cluster analysis (Guerrero et al., 2023).

Annotation efficiency is another area where many existing software solutions fall short. Take Raven Pro (Charif et al., 2010) as an example; it allows detailed annotations and can support long recordings (e.g., 2 h) for extensive analysis. However, it requires users to manually seek vocalisations, which can be highly time-consuming for large-scale datasets. Although beneficial for small datasets, this granularity proves inefficient for large-scale analyses where broader compression and faster processing times are needed. Moreover, some software tools, like Sonobat Szewczak (2010) and BatExplorer (BatExplorer, 2020), are specialised for analysing sounds of specific animal groups. While this specialisation allows for detailed analysis for targeted studies, it severely limits the software's applicability to broader, more diverse natural soundscapes. Furthermore, existing approaches often lack the comprehensiveness needed to capture the full complexity of natural soundscapes, often limiting their focus to specific taxonomic groups or geographic regions (Salamon et al., 2016; Lasseck, 2019; LeBien et al., 2020). This results in models poorly suited for generalisation across different ecosystems.

Many of these limitations stem from gaps in the core features of existing tools. As shown in Table 1, a comprehensive ecoacoustics software tool should address several key capabilities. Data handling is critical for managing, processing, and organising extensive datasets efficiently, especially when working with varied file formats (Truskinger et al., 2014). The user interface must be intuitive and accessible to ensure that researchers can easily navigate and utilise the tool's functionalities without extensive training. Audio playback capabilities, such as adjustable speeds, looping, and synchronised spectrogram visualisations, are essential for sound analysis (Solsona-Berga et al., 2020).

Additionally, status updates help users track the progress of data processing tasks, reducing uncertainty during time-intensive operations. Multi-class annotation supports the labelling of overlapping sound sources within a single sample, which is particularly useful for complex soundscapes (Cartwright et al., 2019). For unpacking high-dimensional data, tools that provide high-dimensional visualisation techniques enable users to identify patterns and clusters effectively (Ishibashi et al., 2020). Advanced segmentation is also assists the division of recordings into meaningful sound events through automated or semi-automated methods.

For users needing flexibility, user-defined algorithms allow customisation of workflows to suit specific research objectives. Flexible environment support ensures compatibility with different operating systems and research setups, broadening the tool's applicability. Finally, efficiency-optimised labelling minimises the effort required for annotations by integrating features like batch labelling and majority label propagation, which are needed for saving time and effort when annotating large-scale datasets.

As shown in Table 1, many existing audio annotation and visualisation systems have limitations in such specialised features. The same applies to software like Sonic Visualiser (Cannam et al., 2010) and Wavesurfer (Sjölander and Beskow, 2000). These general audio editing tools fall short of providing the specialised functionalities required for a comprehensive PAM workflow, such as multi-class annotation and advanced segmentation. Even purchased solutions like SASLab Pro (Specht, 2002), while excelling in user interface and basic functionalities, do not offer full support for advanced features like efficiency-optimised labelling. This patchwork landscape leaves gaps for users who require a comprehensive set of features geared towards large-scale ecoacoustics data annotation.

3. Methodology

3.1. Software design and implementation

LEAVES[1] is a client-based application with a web-based interface developed with Python 3.8 and JavaScript, designed for exploring and annotating large natural soundscape datasets for ecoacoustic research. The application leverages a combination of frameworks and libraries to deliver its functionalities effectively. Frameworks such as Flask provide back-end support for web development, enabling server deployment flexibility, while Dash is used to build the interactive web-based user interface. Together, these frameworks ensure a responsive and scalable platform for data annotation and visualisation.

The application also integrates several specialised libraries to handle data processing, visualisation, and machine learning tasks. Pandas (McKinney et al., 2010) and NumPy (Harris et al., 2020) are employed for data manipulation and numerical computations, respectively. Plotly (Shammamah, 2019) powers the creation of interactive visualisations, while Matplotlib (Hunter, 2007) is used for plotting and visualising audio features and clustering results. For audio analysis, the application utilises librosa (McFee et al., 2015) for feature extraction, PyDub (Robert et al., 2018) for manipulating audio segments, and PyGame (Shinners, 2011) for audio playback functionalities, ensuring an integrated experience for users to listen to and annotate audio samples. Machine learning and clustering tasks are facilitated by Scikit-learn (Pedregosa et al., 2011), which provides algorithms for preprocessing (MinMaxScaler), dimensionality reduction (UMAP (McInnes et al., 2018)), and clustering (DBSCAN).

The decision to store data locally on the user's machine ensures rapid data access and supports a cost-effective infrastructure while maintaining the tool's flexibility to be deployed on a server if needed. This local-first approach gives researchers direct control over their data and ensures faster processing by handling all computations locally on the user's machine. By avoiding the need for server-based processing, we are able to make LEAVES accessible to users with limited internet bandwidth or budget constraints. This also enables it to be used in remote field settings or by research groups with limited access to cloud services.

As such, LEAVES addresses challenges faced by existing software by emphasising real-time feedback, an intuitive user interface, high-quality audio playback, and efficient status updates. These features streamline workflows for large datasets, facilitating detailed auditory analysis while ensuring accuracy through expert validation. As such, it functions as a decision support system, relying on a "human-in-the-loop" approach (Mosqueira-Rey et al., 2023) that ensures robustness and accessibility for diverse research needs.

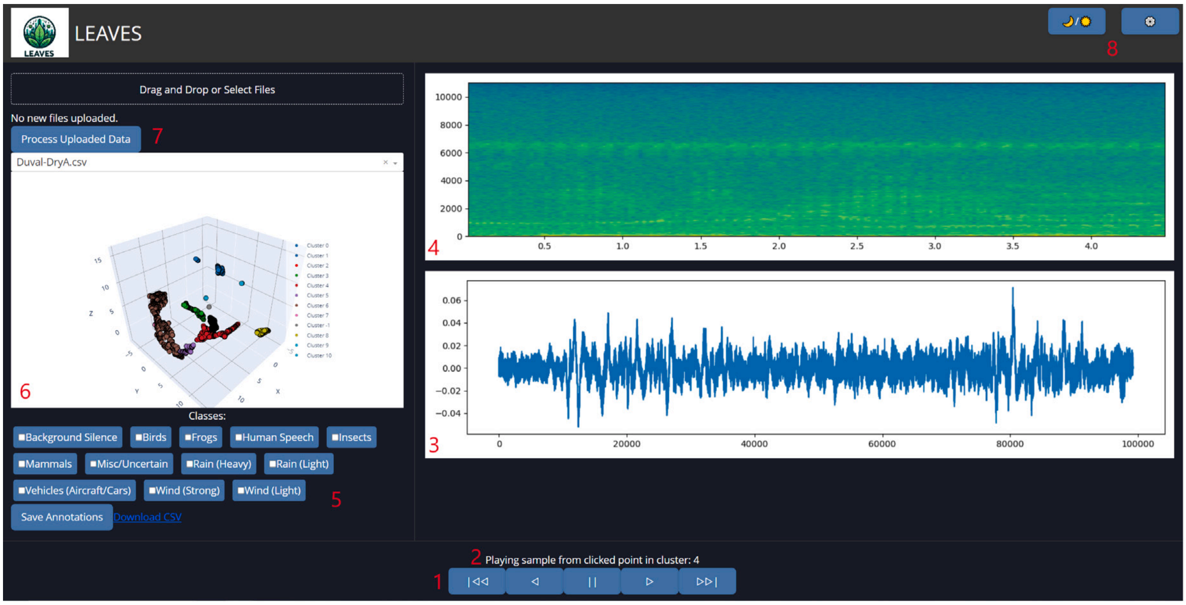
LEAVES is also equipped with advanced functionalities tailored specifically for large-scale ecoacoustics research. Among these is multi-class annotation, allowing users to annotate sounds within a single audio segment across different taxonomic groups and abiotic acoustic events. Although the identification of specific species is important, we designed the system to handle a wider range of ecoacoustic data, including other taxonomic groups such as mammals, amphibians, and insects, as well as environmental and anthropogenic sounds. Several automated approaches, such as BirdNET (Kahl et al., 2021), already specialise in bird species identification. However, there is a gap in tools that address non-avian sounds and mixed soundscapes. By differentiating biophony, geophony, and anthrophony, we can address the need for software that can annotate and analyse the full diversity of sounds present in ecoacoustic datasets, beyond individual species alone (Barber et al., 2011; Grinfeder et al., 2022).

3.2. Workflow design

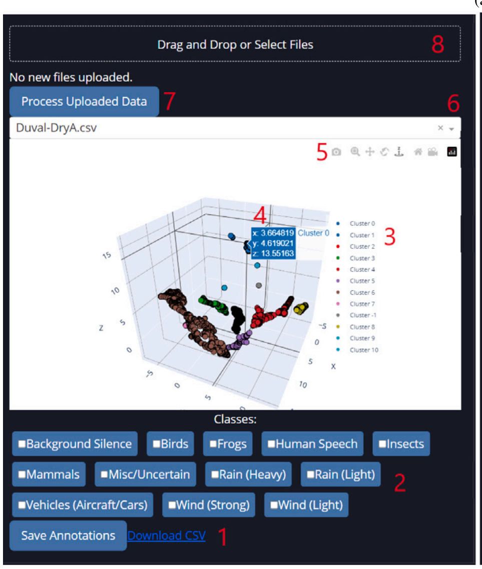
The software enables intuitive uploading, visualisation, and annotation of large-scale ecoacoustic data, using a flexible ingestion module compatible with common environmental sound formats, as seen in Figs. 2(a) and 2(b), which aligns with the data formats common to large-scale ecoacoustics projects like the A2O (Roe et al., 2021). Batch processing and optimised algorithms enable LEAVES to efficiently analyse diverse ecoacoustic sounds, supporting high-throughput clustering and pattern recognition. These features allow simultaneous analysis of multiple files, facilitating faster data handling and effective pattern recognition across large datasets.

To ensure accurate annotations, it is essential to have clear and normalised audio signals. During pre-processing, we apply signal processing techniques to prepare the data for detailed analysis. Users are

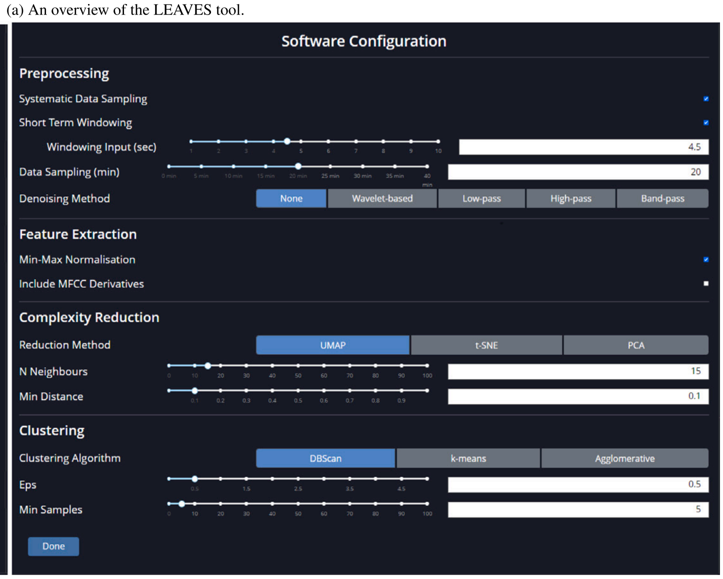
[1] https://github.com/thomasnapier/LEAVES.



(a) An overview of the LEAVES tool.



(b) The data ingestion, 3D visualisation, and annotation interface.



(c) The software configuration modal.

Fig. 2. A breakdown of the LEAVES interface. The numbers correspond to the following sections. (a) 1: The media playback controls including going to previous and next sample as well as skipping to the previous and next cluster. 2: The status label, which is used for showcasing the current sample being annotated as well as displaying other information to the user. 3: The real-time Mel-spectrogram generated per-sample. 4: The real-time waveform plot generated per sample. 5: The annotation classes. 6: The interactive 3D clustering scatterplot with point-click capabilities. 7: The data ingestion module. 8: The day/night mode and settings buttons. (b) 1: The save button and download link for exporting a CSV file with the current annotations. 2: The annotation classes (multiple can be selected per sample). 3: The interactive 3D clustering scatterplot legend. 4: The data-point on-hover information panel. 5: The 3D embedding controls including zoom, orbit, pan and rotate movement options. 6: The file selection drop-down when the data from multiple days or sensors are uploaded. 7. The process uploaded data button. 8. The data upload component which allows users to drag and drop or navigate through their system files. (c) The software configuration modal for changing aspects of the pre-processing, feature extraction, complexity reduction or clustering according to user needs.

then guided to the software configuration interface, where they can fine-tune various audio processing algorithms, such as noise reduction, feature extraction modifiers, and complexity reduction adjustments, as shown in Fig. 2(c). This interface is designed with low entry barriers for new users while providing advanced customisation options for experienced researchers. Following this, the software automatically populates the interactive 3D embedding visualisation, which serves as the control centre for interacting with the audio data, as showcased in Fig. 2(b). Users may interact with any point in any cluster to listen to the associated sample. Alternatively, users may click through a proportion of samples randomly selected from each cluster to label. In either case, the audio sample's associated Mel-spectrogram and waveform are generated in real-time. Keyboard shortcuts are also available to streamline navigation and playback, with arrow keys for moving between samples and the space-bar for starting and stopping playback. Although a playback cursor is not currently implemented, it is planned for implementation in future updates as part of ongoing usability enhancements.

Within the annotation section, illustrated in Fig. 2(b), several customisable annotation tags are designed for multi-class labelling. The design intent behind this is that a user may "select all that apply" for any given audio sample. Each annotation is saved to a locally generated Comma-Separated Values (CSV) file, which can be exported anytime. Once all of the randomly selected samples are annotated by a user, the majority label is propagated to the remaining cluster. If there are an equal number of annotations in multiple classes, a custom tie-breaker

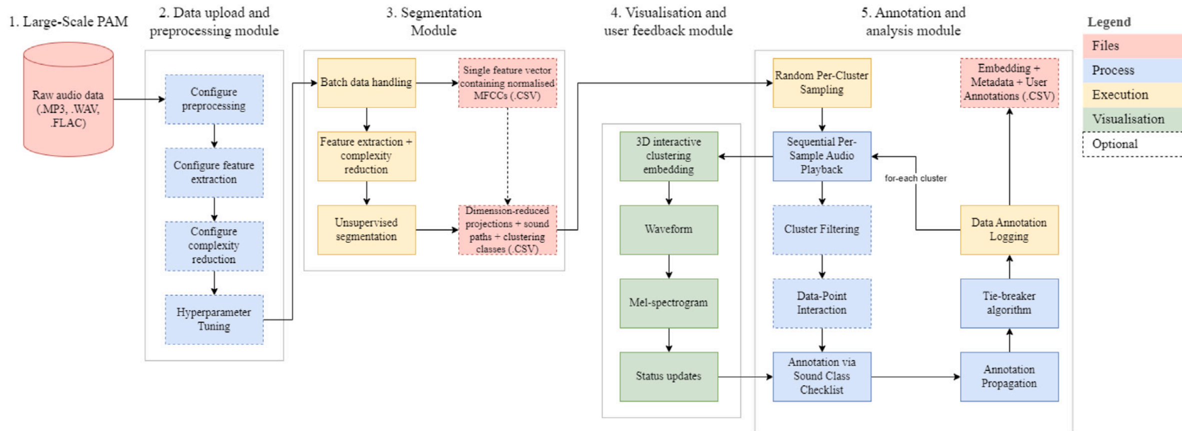


Fig. 3. The processing and algorithm framework of LEAVES. The five main processing steps are described in the corresponding sections of the article.

algorithm determines which annotation propagates to the rest of the cluster. Finally, when users conclude their annotations or save them manually, the tool finalises the annotations, associated metadata, and records the origin of each annotation, distinguishing between manually provided labels and those inferred by the system, ensuring transparency and traceability for downstream analysis.

Balancing the need for granular annotations with the necessity for efficient, high-throughput data processing was a significant challenge. Our solution was to incorporate an efficiency-optimised annotation algorithm that enables users to perform batch annotations on a user-defined proportion of random samples from each cluster. This was achieved by first developing and refining our clustering approach. The tool implements a multi-phase random cluster sampling where data is organised into groups, and subsequent groups are randomly selected, followed by a random selection of members within these groups. This method increases precision, reduces costs, and minimises non-response (Baltes and Ralph, 2020). The tool allows users to adjust the sampling proportion as needed, with a default setting that annotates a 10% sample from each cluster. This feature ensures that researchers can tailor the sampling strategy to their specific project needs, optimising the balance between annotation depth and workload.

3.3. Processes and algorithms

The first module of our system manages ecoacoustic data acquisition and pre-processing to support feature extraction and analysis. It efficiently ingests and normalises diverse ecoacoustics datasets, including large-scale resources like A2O and NEPAN, as well as bespoke collections. Thus, the data input mechanism efficiently ingests and normalises diverse audio recordings, compatible with downstream processes to meet various ecoacoustic research demands, regardless of regional or species specificity. As illustrated in Fig. 3, the data input mechanism is designed to accommodate standard audio formats prevalent in the ecoacoustics domain, including .WAV, .MP3, and .FLAC. We expect batches of raw audio data to be uploaded as a series of continuous recordings, as this is a key characteristic of large-scale ecoacoustics datasets captured by PAM.

After upload, users can adjust pre-processing, sampling, and feature extraction settings. By default, the system samples the first 20 min of audio per hour, as this approach aligns with certain ecoacoustic practices from a similar region (Linke et al., 2018). However, we made this a parameter that can be fine-tuned by the user to meet the specific requirements of their study during the pre-processing configuration stage. Input signals are divided into adjustable temporal segments, defaulting to 4.5-s non-overlapping intervals to capture both short and extended vocalisations efficiently. The optional denoising module includes scalable filters, such as wavelet-based, low-, high-, and band-pass, which are known for their scalability and widespread usage in ecoacoustics for enhancing the clarity of the audio data across a range of applications (Alonso et al., 2017; Brown et al., 2018).

Feature extraction is a critical phase, where Mel-spectrograms with 128 Mel bands are generated for visualisation alongside the computation of Mel-Frequency Cepstral Coefficients (MFCCs) and their derivatives. This dual approach facilitates easier annotation through visual cues and harnesses MFCCs' proven efficacy in encapsulating the intricate details of acoustic signals (Mesaros et al., 2018; Mcloughlin et al., 2019). The subsequent concatenation into a singular feature vector and the application of min–max normalisation ensures a uniform scale across all features, mitigating any disproportionate influence from outlier data points. To manage data dimensionality, we implemented Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018), t-distributed Stochastic Neighbour Embedding (t-SNE) (Van der Maaten and Hinton, 2008) and Principal Component Analysis (PCA) (Maćkiewicz and Ratajczak, 1993), which are widely used in scientific and ecoacoustic analysis (Huang et al., 2019). Upon selection of PCA, cumulative explained variance ratio is used as a determinant for selecting the optimal number of components. Both t-SNE and UMAP, however, are sensitive to hyperparameters, namely perplexity as well as number of neighbours and minimum distance, respectively, and thus, we allow users the option to tune them. However, we note that our default selection is UMAP due to its well-documented mathematical foundation and improved clustering accuracy (Allaoui et al., 2020).

The tool implements several popular clustering algorithms for unsupervised segmentation, including k-means (MacQueen et al., 1967) and DBSCAN (Schubert et al., 2017). By default, our system has been designed to leverage the DBSCAN algorithm for its adaptability and resilience to noise, which aligns well with the characteristics of ecological audio data. The segmentation process can be fine-tuned through hyperparameter adjustments, including n-neighbours and $min\text{-}dist$, alongside exploring various distance metrics.

The visualisation component includes a 3D embedding viewer with interactive features like zoom, pan, rotate, and orbit, allowing complete model inspection. The interface highlights clusters during annotation for focused analysis and supports filtering by cluster attributes. To enhance usability, we integrated Mel-spectrogram and waveform displays with the 3D viewer, enabling intuitive navigation through samples with media controls for efficient, user-friendly annotation. The random per-cluster sampling method selects a user-defined subset of samples from each cluster, with a default of 10%. This strategy ensures representative annotations in large datasets. After annotating the subset, the majority label is applied to the entire cluster. In cases of a tie, a tie-breaker algorithm uses a predefined label hierarchy, allowing users to make the final selection based on domain knowledge.

The final CSV file generated is designed for maximum interoperability, following proposed metadata standards for annotations (Roch et al., 2016; Vella et al., 2022). These standards ensure that the data can be easily integrated with other tools and databases, promoting data sharing and collaboration. Adhering to the FAIR principles, the metadata includes detailed information about the recording environment, including the geographic location, recorder name, date and time; the annotations, including a 0 or 1 in place representing the absence or presence of each sound class for each given sample, and the specific x, y and z embedding coordinates of each sample to enable re-upload to the tool if needed. This approach aligns with the best practices in open science, facilitating transparency and reproducibility in ecoacoustic research.

4. Application to a continent-scale study of ecoacoustic monitoring

4.1. Software

The developed software tool integrates the core functionalities described in Section 2.1, executes locally, and can accept up to several weeks of data at once, which can be paused, resumed and even recovered if an upload is interrupted, without losing state. It notably omits cloud resource management and offers a basic but functional input interface, only designed and intended to aid in research.

The tool employs a data-locality scheduling approach to optimise processing, with Python and associated libraries, including Librosa, Plotly, and Dash, supporting the workflow framework and processing of underlying ecoacoustic samples. With this initial design, we emphasised efficiency, scalability, and adaptability, setting a foundation for future enhancements.

4.2. Testing methodology

The testing methodology focused on validating the clustering and annotation process through internal and external clustering evaluations. Internal validation assessed cluster quality, while external validation evaluated cluster accuracy with real-world ground truth labels. Additionally, we conducted a scalability assessment to measure the time-saving potential of the software-assisted approach.

We selected six sites from the A2O network spanning the eastern side of Australia, which are currently under close investigation by ecologists (Allen-Ankins et al., 2023). These sites were chosen due to their ecological significance and the variety of soundscapes they present, featuring a broad spectrum of species and acoustic signatures representative of various ecoregions. We systematically extracted the first 20 min from each 2-h segment in line with the literature (Linke and Deretic, 2019) and, as advised by domain experts, amounting to 4 h of data per recorder at each site. Data was then processed through our workflow using the tool's default parameters: segments were split into 4.5-s non-overlapping parts without applying denoising techniques. MFCCs were extracted from each slice, with their derivatives excluded. Subsequently, the data underwent dimensionality reduction via UMAP, followed by two clustering approaches: DBSCAN and k-means.

It is crucial to note that UMAP, DBSCAN, and k-means are all hyperparameter-sensitive algorithms, meaning that the selection of these parameters can drastically change the final result. To evaluate the internal cluster validity, we performed an extensive grid-search analysis to find the optimal values for each of these according to the performance scores in Table 2. For external validation, we manually labelled a single cluster from each site, with cluster sizes ranging from 30 to 223 points. During this, we observed that each individual sample typically included no more than two dominant sound sources at any one time. A dominant sound source, in this context, does not imply exclusivity but indicates the primary sources that are most audible. For instance, while insect sounds were present in many recordings, they were not always the dominant source, particularly when sounds like birds and rain were also present (Phillips et al., 2018). This observation aligns with the complexity and richness of ecoacoustic environments where multiple sound sources coexist.

Furthermore, we implemented a nested labelling approach to address the complexity and hierarchical nature of ecoacoustic data. This approach considers partial matches and nested relationships between labels, providing a more accurate measure of clustering performance. For example, if a cluster contains segments labelled as *insects + birds* and *insects*, our nested labelling approach recognises that *insects* is a subset and partial match of *insects + birds*. By acknowledging this hierarchical relationship between labels, we ensured that even the intricate and overlapping sound sources were accurately represented in our analysis.

4.3. Scalability assessment

We conducted a small test using one cluster from each site during manual labelling, equating to 882 points total. Each 4.5-s sample was manually labelled, and the time taken for this task was recorded. To ensure independence between the software-assisted and manual labelling processes, the individual conducting the manual labelling was not informed of the cluster assignments or labels generated during the process. Instead, they were provided only with a list of sound files from the selected clusters, with no reference to the software outputs. This separation of information minimised the influence of prior exposure to the data, thereby reducing potential bias in the manual labelling process. Labelling the entire subset of clusters took approximately 166 min, resulting in an average time of 11.3 s per sample. This duration corresponds to about 2–3 listens per sample, or a listening ratio of 2.51. We compared the software-assisted labelling approach with manual labelling across six A2O datasets by visualising the cumulative labelling progress over time, with manual labelling time modelled as a linear progression based on the average time taken.

In contrast, the software-assisted method was predicated on the premise that only a 10% subset within each cluster required manual annotation, although this number could be modified for more fine-grained annotations. This choice is supported by empirical evidence and statistical principles. Recent research on clustering algorithms recommends aiming for sample sizes of $N = 20$ to $N = 30$ per expected subgroup for satisfactory power, particularly when large subgroup separation is anticipated (Dalmaijer et al., 2022). In our study, the 10% sample size typically falls within or exceeds this range, ensuring sufficient samples for accurate analysis. Furthermore, our empirical observations confirm that annotating 10% of the samples in each cluster was generally sufficient to identify a similar number of unique sound classes, supporting the effectiveness of this sample size for achieving comprehensive coverage of sound diversity.

After this initial manual labelling phase, labels are extrapolated across the entirety of each cluster, introducing a non-linear progression in the cumulative percentage of the labelled dataset. This approach necessitated a detailed examination of specific cluster sizes within each dataset, ranging from clusters containing only 7 points up to large clusters of 1426 samples. Our analytical framework aimed to interpolate the labelling progression across these variable cluster sizes, synthesising a smooth, average curve to represent the cumulative percentage of the dataset labelled over time. This was achieved by simulating 10 runs for each cluster configuration and interpolating the resulting labelling curves to a set of common time points. The average of these interpolated values at each time point were then calculated to yield the average curve for each dataset.

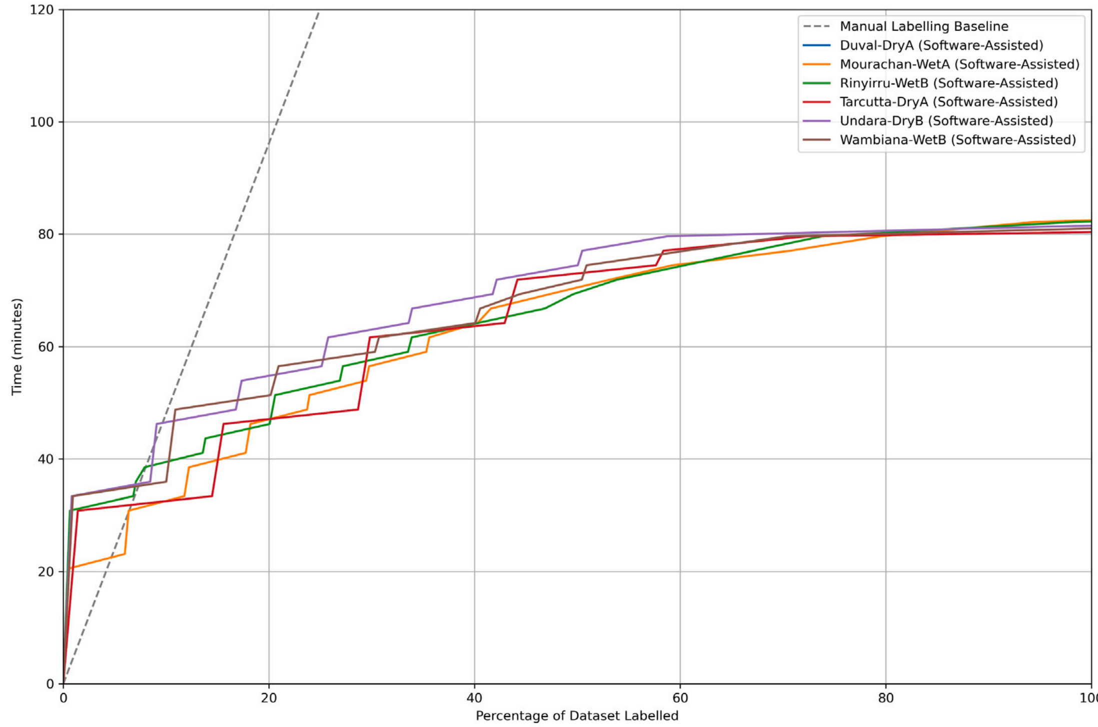


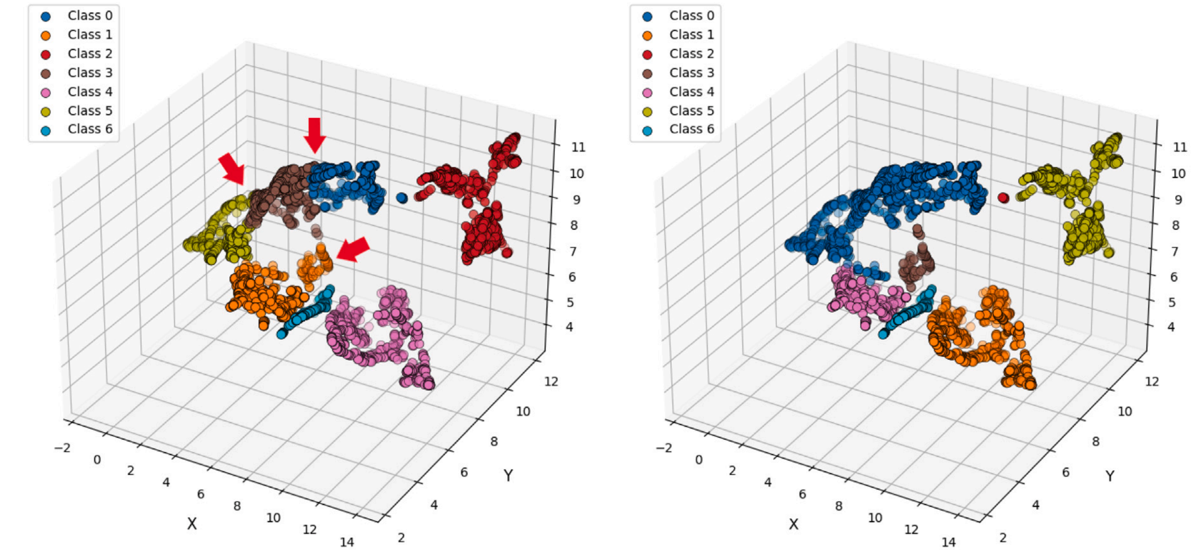
Fig. 4. A comparison of labelling efficiency showing the cumulative percentage of datasets labelled over time. The manual labelling line serves as a baseline, with an average time required of 11.3 s per 4.5-s sample. In contrast, the software-assisted approaches demonstrates significant efficiency gains, particularly for larger clusters, highlighting its potential for scalable data annotation.

Table 2
Mathematical formulas and definitions for internal and external clustering evaluation metrics.

Metric	Formula	Definition	Reference
Internal metrics			
Silhouette Coefficient	$S = \frac{b-a}{\max(a,b)}$	Measures how similar an object is to its own cluster compared to other clusters.	Rousseeuw (1987)
Calinski-Harabasz index	$CH = \frac{BCSS/(k-1)}{WCSS/(n-k)}$	Evaluates clusters based on the ratio of between-cluster dispersion to within-cluster dispersion.	Caliński and Harabasz (1974)
Davies–Bouldin index	$DB = \frac{1}{n}\sum_{i=1}^{n}\max_{i\neq j}\left(\frac{\sigma_i+\sigma_j}{d(c_i,c_j)}\right)$	Assesses the average similarity between each cluster and its most similar one.	Davies and Bouldin (1979)
Dunn index	$DI = \min_{1\leq i<j\leq n}\left(\frac{\delta(c_i,c_j)}{\max_{1\leq k\leq n}\Delta(c_k)}\right)$	Identifies compact and well-separated clusters, measuring the ratio of the smallest inter-cluster distance to the largest intra-cluster distance.	Dunn (1974)
Composite Score	$CS = (N_S)^2 + (N_{CH})^2 + (1-N_{DB})^2 + (N_{DI})^2$, where $N_m = \frac{m-\min(m)}{\max(m)-\min(m)}$	Aggregates normalised and squared metric values to provide a comprehensive clustering quality measure. N_m denotes the normalised value of metric m.	
External metrics			
Adjusted Rand index	$ARI = \frac{RI - \text{Expected RI}}{\text{Max RI} - \text{Expected RI}}$	Measures the similarity between two data clusterings, adjusted for chance.	Hubert and Arabie (1985)
Normalised Mutual Information	$NMI = \frac{2\cdot I(X;Y)}{H(X)+H(Y)}$	Evaluates the mutual dependence between the clustering and the ground truth.	Strehl and Ghosh (2002)
Fowlkes-Mallows index	$FMI = \sqrt{\frac{TP}{TP+FP}\cdot\frac{TP}{TP+FN}}$	Assesses the similarity between two clusterings based on precision and recall.	Fowlkes and Mallows (1983)

4.4. Results and evaluation

Our results demonstrate the efficiency of the software-assisted labelling approach, particularly after initial manual labelling of each cluster. Fig. 4 shows a non-linear increase in dataset labelling over time, contrasting with the linear progression of manual labelling and highlighting the accelerated capability of the software-assisted approach. The derived average curves for each dataset revealed a distinct, non-linear increase in the percentage of the dataset labelled over time, signifying rapid gains once the initial manual annotations were propagated across the remaining cluster points. This pattern is contrasted with the linear progression observed in manual labelling, underscoring the accelerated labelling capability of the software. The systems method's efficiency gains are particularly evident in datasets with



(a) Clustering of ecoacoustics samples using the k-means algorithm. The large clusters in the embedding are split into spherical-shaped clusters, as highlighted by the red arrows, demonstrating the tendency of k-means to form spherical clusters.

(b) Clustering of ecoacoustics samples using the DBSCAN algorithm. Unlike k-means, DBSCAN maintains larger, irregularly shaped clusters, demonstrating its ability to find non-linear boundaries.

Fig. 5. Comparative 3D UMAP scatterplots of ecoacoustic samples from the Duval region, illustrating differences in cluster formation using k-means (a) and DBSCAN (b). k-means splits a large cluster into three distinct sub-clusters that better represent the underlying sound types. In contrast, DBSCAN combines these into a single cluster, which, while preserving the broader structure, lacks the granularity provided by k-means. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

larger clusters, where manually labelling a small subset effectively propagates to larger data groups. The early intersection of the software-assisted curve with the manual labelling trajectory emphasises the scalability of the approach, with time savings increasing proportionally to dataset size. For example, at 10% of the dataset labelled, the software-assisted approach already shows significant time savings compared to manual labelling.

From these results, we calculated that manual approaches require approximately 605.8 min to annotate 3205 4.5-s-long samples, whereas our method takes only 85 min while achieving high internal clustering accuracy. This represents a significant reduction in time, indicating that the software-assisted is approximately 7.12 times faster than the manual method. By grouping sounds holistically—capturing a wide range of audio classes beyond individual species—the tool provides a broad ecological overview that streamlines initial data processing. As such, the quantity of validation required will vary depending on the study's objectives, with annotations closer to species-levels necessitating a higher time investment. However, the algorithms and workflows integrated into the tool, such as UMAP and clustering approaches like DBSCAN, are well-suited for such applications. Previous studies, such as the clustering of neotropical passerines based on vocalisations (e.g., rough-legged tyrannulet, Phyllomyias burmeisteri), demonstrate the feasibility of leveraging these techniques for species-level segmentation (Parra-Hernández et al., 2020).

4.5. Clustering evaluation metrics

Clustering performance without ground truth labels was evaluated using metrics that quantify clustering quality based on dataset properties (Liu et al., 2013). Table 2 presents these metrics and a new composite score, with each metric selected to offer unique insights into clustering results. Below, we provide our justifications for the selection of each metric:

- **Silhouette Coefficient**: This metric gauges how similar an object is to its own cluster compared to others. The Silhouette Coefficient's value ranges from −1 (incorrect clustering) to +1 (highly dense clustering), with values near zero suggesting overlapping clusters. While it provides a clear measure of cluster tightness and separation, its preference for convex clusters might not fully represent the quality of clusters formed by algorithms like DBSCAN, which can identify arbitrary-shaped clusters.
- **Calinski-Harabasz Index**: Also known as the Variance Ratio Criterion, this index measures the ratio of the sum of between-cluster dispersion to within-cluster dispersion. It is effective in identifying dense and well-separated clusters, which is ideal for convex cluster shapes but may not be as informative for clusters of varying densities and non-spherical shapes.
- **Davies–Bouldin Index**: This index evaluates the average similarity between each cluster and its most similar one, where lower scores denote better separation. The Davies–Bouldin Index is favoured for its simplicity and computational efficiency. However, like the Silhouette Coefficient, it relies on Euclidean distances, which might not suit all types of cluster structures.
- **Dunn Index**: We specifically selected the Dunn Index for its capacity to identify sets of compact and well-separated clusters, irrespective of their shape. This index is particularly useful for algorithms like DBSCAN that can produce non-convex clusters. It compares the smallest distance between observations in different clusters to the largest intra-cluster distance, with higher values indicating better clustering.

Furthering this, we formulated a new composite score to synthesise the insights provided by individual internal metrics into a single evaluation measure. The score consists of aggregated insights from individual clustering metrics to evaluate cluster compactness, separation, and dispersion more holistically. Specifically, the Silhouette Coefficient focuses on within-cluster tightness relative to inter-cluster distances, while the Dunn Index excels in identifying compact and well-separated clusters regardless of shape. The Calinski-Harabasz Index measures the variance ratio, which is ideal for assessing dense and convex clusters, and the Davies–Bouldin Index evaluates cluster separation based on similarity.

By normalising and squaring the contributions of each metric, the composite score ensures that no single metric disproportionately influences the evaluation, providing an overview of clustering quality. In practice, the composite score aids in identifying clustering results

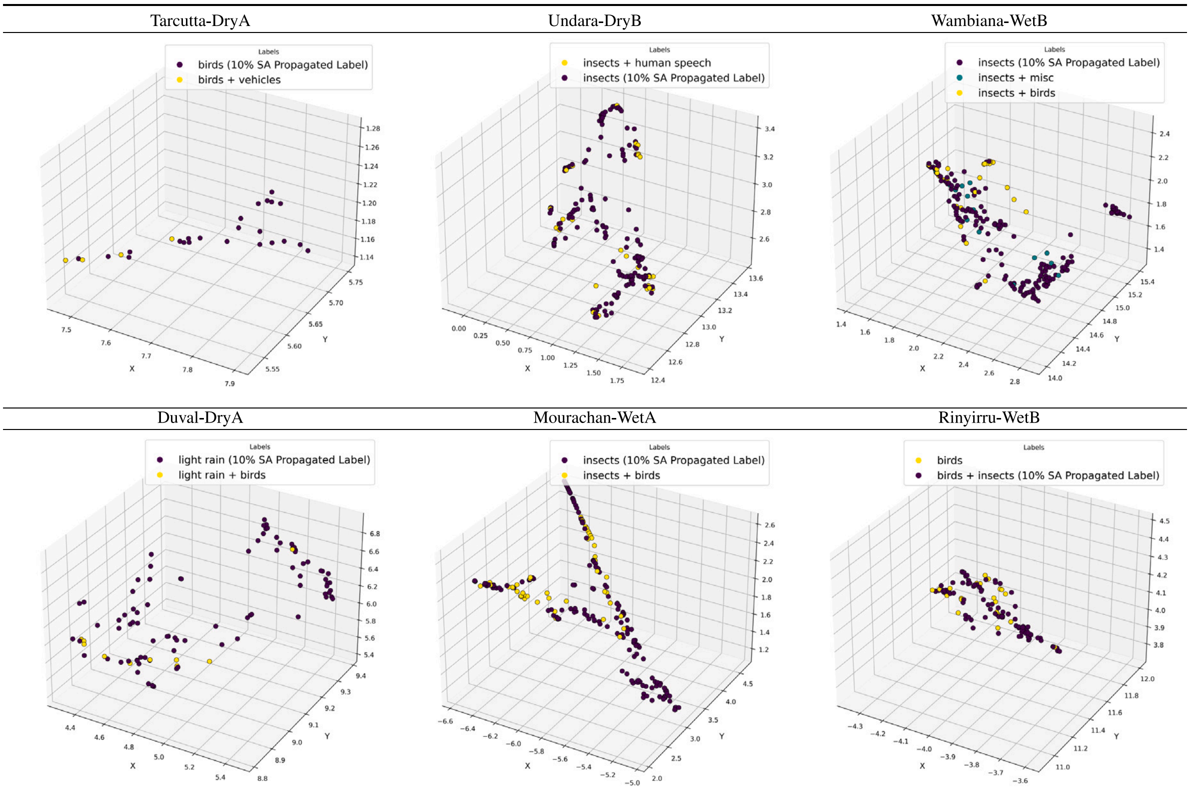


Fig. 6. Grid of embedding subplots for six different sites, showing UMAP and DBSCAN 3D scatterplots with ground truth labels overlayed. The legend indicates the ground truth labels (yellow), with matching ground truth label returned by the 10% labelling software-assisted approach marked (purple). This demonstrates that the 10% labelling correctly identified the most common sound-class in each cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

that optimise compactness and separation, offering a straightforward decision-making tool to compare different clustering configurations. The original sub-indices remain available for detailed analyses when needed.

On the other hand, assessing clustering performance with ground truth labels requires metrics that can accurately quantify the alignment between the predicted clustering and the known labels. These metrics collectively offer a comprehensive view of the clustering performance, accommodating different cluster shapes and densities.

- **Adjusted Rand Index**: This measures the similarity between two data clusterings, adjusting for chance. It accounts for both True Positive (TP) and True Negative (TN) decisions, making it robust for different cluster sizes and numbers. High values indicate strong agreement with the ground truth. It is particularly useful in scenarios where the number of clusters in the ground truth may vary significantly from the clustering result.
- **Normalised Mutual Information**: This evaluates the mutual dependence between the clustering and the ground truth. It scales the mutual information score to a range between 0 and 1, facilitating comparison across different datasets. High values suggest that the clustering captures a significant amount of the information present in the ground truth. This metric is valuable for its ability to handle different clustering sizes and is invariant to permutations of the cluster labels.
- **Fowlkes-Mallows Index**: This assesses the similarity between two clusterings based on precision and recall, offering a balanced measure of clustering quality. It is particularly useful for datasets where the focus is on the accurate identification of pairs of points that belong to the same cluster. High values indicate that the clustering algorithm has a high TP rate and low False Positive (FP) and False Negative (FN) rates. This metric is advantageous when the interest is in the precise recovery of the cluster structure. Notably, the FMI can be calculated for both hierarchical and non-hierarchical clustering results, providing a comprehensive method for evaluating clustering performance across different approaches. This is possible because the FMI relies on pairwise comparisons of data points rather than on the overall structure of the clustering, making it adaptable to different clustering outputs.

4.6. Clustering performance results

4.6.1. Internal clustering performance

The analysis of the clustering performance of our software tool utilising both DBSCAN and k-means algorithms demonstrates a strong ability to delineate well-separated, distinct clusters across various complex datasets. Fig. 5 focuses on the Duval region, comparing k-means (Fig. 5(a)) and DBSCAN (Fig. 5(b)) clustering. This figure illustrates that large clusters in the embedding are split into spherical-shaped clusters when clustered using k-means (highlighted with red arrows in Fig. 5(a)), whereas DBSCAN does not produce this effect. However, looking at Table A.4, it is evident that k-means performs better for these types of datasets due to the higher composite score it achieves. In datasets with more defined and spherical cluster distributions, such as Duval-DryA and Tarcutta-DryA, k-means performed well, particularly in the Silhouette and Calinski-Harabasz scores. The consistent performance of k-means in these scenarios illustrates its effectiveness at enhancing cluster tightness and overall definition, which are important for accurate data analysis and annotation.

As seen in Fig. 6, which displays the DBSCAN scatterplots for all six sites, this assessment, grounded in the application of a comprehensive

Table 3
External validation results for the selected clusters in each dataset.

Dataset	Ground truth labels	Software-assisted propagated label	FMI (Non-hierarchical)	ARI (Hierarchical)	NMI (Hierarchical)	FMI (Hierarchical)	Total labels	Incorrect labels (%)
Tarcutta-DryA	{birds, birds + vehicles}	birds	0.8723	1.0000	1.0000	1.0000	30	4 (13.33%)
Undara-DryB	{insects + human speech, insects}	insects	0.8450	1.0000	1.0000	1.0000	198	34 (17.17%)
Wambiana-WetB	{insects, insects + misc, insects + birds}	insects	0.8301	1.0000	1.0000	1.0000	223	40 (17.94%)
Duval-DryA	{light rain, light rain + birds}	light rain	0.9027	1.0000	1.0000	1.0000	98	10 (10.20%)
Mourachan-WetA	{insects, insects + birds}	insects	0.7987	1.0000	1.0000	1.0000	212	50 (23.58%)
Rinyirru-WetB	{birds, birds + insects}	birds + insects	0.8123	1.0000	1.0000	1.0000	121	26 (21.48%)

suite of evaluation metrics, shows the tool's capability to reveal the underlying cluster structures inherent to the data. DBSCAN's performance is particularly effective in managing datasets with complex, non-linear boundaries or varied densities. For example, in the Wambiana-WetB dataset, DBSCAN achieved the highest composite score, indicating its effectiveness in extracting well-separated clusters even in challenging environments.

This is further supported by its consistently high scores across the Dunn Index, which measures the extent of separation between the closest clusters. The Mourachan-WetA and Rinyirru-WetB datasets also demonstrated strong clustering performance under the DBSCAN method, which scored higher in both the Silhouette Coefficient and Davies–Bouldin Index, indicative of well-separated and compact clusters. These metrics imply that DBSCAN can identify and enhance the natural cluster structures without being constrained by the tendency to form spherical clusters, as is often the case with k-means.

The findings from this analysis indicate that the software tool is adaptable to different data characteristics. The tool consistently extracts meaningful, well-separated clusters in both simple, spherical clusters and complex, irregularly shaped groupings. As such, the results from this evaluation show that our tool has the ability to operate across various scenarios—from highly structured to highly complex datasets and deliver clear, actionable clustering outcomes.

4.6.2. External clustering performance

The results in Table 3 show that our software-assisted propagation method achieved high FMI values, indicating a strong agreement between software-assisted and ground truth labels. For instance, the FMI for Tarcutta-DryA was 0.8723, and for Duval-DryA, it reached 0.9027, demonstrating the method's robustness in correctly propagating labels within clusters. The hierarchical metrics (ARI, NMI, and FMI) all reached perfect scores (1.0000), reflecting the accuracy of the nested labelling approach in handling complex and overlapping sound sources.

In addition to the metrics used for evaluating clustering accuracy (e.g., FMI, ARI, and NMI), we quantified the number of incorrect labels propagated by the software-assisted method. Table 3 shows the total number of labels, the number of incorrect labels, and the corresponding percentage for each dataset. Across all datasets, the percentage of incorrect labels ranged from 10.20% to 23.58%, with an average error rate of 17.12%. These results demonstrate that the software-assisted approach not only accelerates the annotation process but also maintains a reasonable level of accuracy suitable for large-scale ecoacoustic studies.

By manually validating 10% of each cluster and propagating these labels, we demonstrate that the system can quickly and accurately annotate large-scale datasets, maintaining high levels of correctness and efficiency, as illustrated in Fig. 6. The high agreement between manual and propagated labels confirms the practical utility of our approach in large-scale ecoacoustic data annotation.

5. Discussion

5.1. Impact and applications

The LEAVES tool is designed as a pre-filtering, high-level clustering platform that provides an efficient workflow for organising large-scale ecoacoustic data into coarse label groups, such as "bird", "insect", "rain", "vehicles". Unlike tools focused on species-specific or individual-level identification, LEAVES clusters are intended to facilitate the initial organisation and annotation of vast, unstructured datasets. This approach leverages clustering for general, ecologically meaningful categorisations, creating a flexible starting point for ecoacoustics research that does not necessitate detailed tuning or exhaustive labelling at the species level. Such clustering aligns with goals in ecoacoustics of assessing broad ecological patterns, which often precedes detailed taxonomic identification and offers practical efficiencies for long-term, large-scale data analysis (Gibb et al., 2018; Farina and Gage, 2017).

By providing coarse taxonomic groupings, LEAVES directly supports ecoacoustic studies focused on soundscape composition and the ecological impacts of anthropogenic change. Broad categorisation into groups such as biophony, geophony, and anthrophony facilitates quick insights into habitat composition, biodiversity patterns, and environmental changes over time, enabling ecologists to examine soundscape trends without an immediate need for fine-grained, species-specific data. For instance, studies monitoring the effects of road noise on wildlife can initially classify sound data by general groups, such as separating "bird" from "vehicle" clusters, to observe patterns in wildlife activity in response to human-made sounds (Pijanowski et al., 2011; Merchant et al., 2015). Additionally, clustering can identify seasonal shifts in biophonic dominance — such as insects versus birds — allowing researchers to filter data more effectively for further study of particular taxa or seasonal changes (Linke and Deretic, 2019).

The clustering provided by LEAVES offers practical workflow improvements, markedly reducing the time and effort needed for manual data annotation in large datasets. This clustering approach allows researchers to quickly filter sound classes and focus on relevant subsets, supporting studies that require processing of extensive temporal and geographic data. Unlike manual annotation or highly specific identification methods, LEAVES' general clustering enables broad-scale ecoacoustic studies to be conducted efficiently, conserving valuable researcher time and extending the scope of ecoacoustic analyses (Stowell and Sueur, 2020).

Although LEAVES is not presently aimed at achieving species-level clustering, this general filtering step significantly aids in establishing datasets that can later be refined through targeted approaches or with additional taxonomic features. Potential future developments, such as integrating more detailed acoustic indices or hierarchical clustering, may support higher-resolution analysis if species-level distinctions become necessary. However, LEAVES' primary contribution lies in its

Table A.4
Comparison of internal clustering performance metrics for DBSCAN and *k*-means methods across various datasets, with a composite score summarising overall clustering quality. Metrics are categorised into Cluster Distinction (Between-to-Within Dispersion (Calinski-Harabasz) and Inter-to-Intra Cluster Distance Ratio (Dunn Index)) and Cluster Quality (Silhouette Score and Cluster Separation Score (Davies–Bouldin)). The best-performing results for each metric and dataset are highlighted.

Dataset	Cluster distinction metrics		Cluster quality metrics		Composite score
	Calinski-Harabasz	Dunn index	Silhouette	Davies–Bouldin	
DBSCAN Clustering performance					
Duval-DryA	665.01	**0.322**	0.397	0.637	0.1914
Mourachan-WetA	6481.30	**0.417**	**0.733**	**0.359**	**2.3707**
Rinyirru-WetB	4620.01	**0.377**	0.654	**0.324**	**1.8819**
Tarcutta-DryA	3248.48	**0.356**	0.586	0.516	0.7494
Undara-DryB	1564.99	**0.084**	0.475	**0.453**	**0.3609**
Wambiana-WetB	10 160.87	**0.628**	0.745	0.302	**4.0000**
k-means Clustering performance					
Duval-DryA	**5778.33**	0.0001	**0.5613**	**0.6094**	**1.0000**
Mourachan-WetA	**6725.93**	0.0000	0.7183	0.3958	0.9289
Rinyirru-WetB	**8438.32**	0.0000	**0.6982**	0.4024	0.8654
Tarcutta-DryA	**8570.86**	0.0000	**0.687**	**0.397**	**0.8329**
Undara-DryB	**5839.83**	0.0000	**0.6300**	0.4580	0.2963
Wambiana-WetB	**14 222.55**	0.0000	**0.7774**	**0.2667**	3.0000

ability to triage large ecoacoustic datasets into meaningful clusters, providing an accessible platform for initial data structuring that enables more targeted studies to follow.

5.2. User needs and accessibility

While the tool currently offers advanced capabilities suited for ecoacoustic experts, ongoing developments aim to make LEAVES accessible to a wider audience, including non-specialists and citizen scientists. This expansion acknowledges the increasing role of community science and public engagement in ecological monitoring, with non-expert participation enhancing data coverage and providing valuable local insights (Gibb et al., 2018).

To bridge the gap for non-expert users, future iterations of LEAVES will include user-centred features such as guided workflows, simplified interface options, and integrated tutorial modules. These features are designed to reduce the learning curve and facilitate intuitive interactions, enabling non-expert users to contribute meaningfully to ecoacoustic studies without requiring deep technical expertise. For example, LEAVES could enable citizen scientists to label basic sound categories (e.g., "birds", "human speech") while leaving more complex annotations for expert review. This tiered annotation approach leverages community involvement while ensuring data quality through expert oversight.

5.3. Ongoing and future developments

As ecoacoustic research continues to evolve, it is essential to adapt and expand the capabilities of LEAVES to meet emerging needs and challenges. This section outlines several key areas of ongoing and future developments aimed at enhancing its flexibility, efficiency, and precision.

- Hierarchical class list structuring: We plan to develop hierarchical class list structuring to support multi-level annotation. This structuring will allow for high-level and low-level annotation at both an ecosystem level (e.g., biophony, geophony, anthrophony) and at lower classification levels (e.g., birds and mammals rather than 'biophony') and individual species. This hierarchical approach will facilitate comprehensive soundscape analysis, enabling detailed studies from broad ecosystem assessments to focused species-specific research.
- Automated hyperparameter tuning: Future versions of LEAVES will allow users to automate the hyperparameter tuning process for sensitive dimension reduction techniques (*t*-SNE, UMAP) and unsupervised learning algorithms (DBSCAN, *k*-means) using a grid search approach. This enhancement will aim to optimise clustering outcomes and dimensional reductions based on the specific characteristics of each dataset, thus improving both the accuracy and efficiency of the annotation process while unburdening non-technical users from having to choose optimal parameter settings.
- Additional feature sets: The integration of additional feature sets, such as acoustic indices, could significantly enhance the clustering and analysis capabilities of the tool. Acoustic indices, which quantify various aspects of the soundscape (e.g., diversity, complexity, and intensity), are often used in the literature. Incorporating these indices could provide more nuanced and comprehensive insights into the ecological data.
- Cloud integration: Cloud integration will be a key development that will provide scalable storage solutions, facilitate collaborative work, and enable advanced computational processes without requiring extensive local computing resources. Cloud capabilities will ensure adaptability to the evolving demands of ecoacoustic research and support the growing trend towards collaborative, distributed research efforts.
- Improved data sharing and cross-verification: To enhance the reliability of annotated data, we plan to implement features that allow easy cross-checking of samples within labelled datasets. This will be particularly useful in collaborative settings where annotations may be performed by less experienced users. Integrating collaboration methods and data-sharing capabilities will foster a more cohesive and accurate ecoacoustic analysis framework.
- Expansion to R: While LEAVES currently leverages Python for scalability and performance on large ecoacoustic datasets, we recognise the importance of R in the ecological research community. Popular R packages like warbleR (Araya-Salas and Smith-Vidaurre, 2017) and Seewave (Sueur et al., 2008) offer valuable packages for sound analysis. In the future, our goal is to explore an R-based implementation or an API integration between Python and R. This dual approach will make LEAVES more accessible to ecologists while maintaining the performance needed to handle large datasets efficiently. This extension will align with our goal of fostering collaboration between disciplines and broadening the user base.

These planned features will make LEAVES more flexible and useful for various ecoacoustics analysis use cases.

6. Conclusion

LEAVES represents a meaningful advancement in ecoacoustic data management, providing an efficient, high-level clustering framework that enables rapid pre-filtering and annotation of large-scale soundscapes. By grouping data into broad taxonomic categories, LEAVES enhances the speed and scalability of soundscape analysis, addressing a critical need in ecoacoustics research for software that can handle extensive datasets with diverse sound sources. This clustering-based approach is essential for studies aiming to understand biodiversity patterns and ecological dynamics, offering a robust foundation for initial data structuring and enabling ecologists to focus on broad soundscape trends before delving into finer species-specific details.

Although LEAVES significantly improves annotation efficiency, it emphasises the importance of human expertise to verify the clustered data, positioning itself as a decision support system that integrates human oversight to maintain annotation reliability. This human-in-the-loop framework balances automation with expert validation, ensuring the production of high-quality data without sacrificing scalability.

This improvement has broad implications for ecological research, offering an alternative open-source tool to assist in understanding biodiversity and ecosystem dynamics. Future directions include further refinement of specific features, such as customisable class lists, automated hyperparameter tuning, and cloud integration, to enhance user experience and efficiency. Additionally, the potential for community contributions, including those from citizen scientists, could greatly expand the tool's applicability and data richness. This collaborative approach and ongoing developments emphasise the software's adaptability to evolving research needs.

CRediT authorship contribution statement

Thomas Napier: Writing – original draft, Visualisation, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualisation. **Euijoon Ahn:** Writing – review & editing, Supervision. **Slade Allen-Ankins:** Writing – review & editing, Supervision. **Lin Schwarzkopf:** Writing – review & editing, Supervision. **Ickjai Lee:** Writing – review & editing, Supervision, Project administration.

Software availability

To download LEAVES, users can visit our GitHub repository at https://github.com/thomasnapier/LEAVES/, which includes installation files, usage instructions, link to a web-hosted demo and all required dependencies. We recommend following the installation guide in the repository for local use of the tool. A step-by-step guide, along with detailed download and installation instructions, is also available on the project page at https://thomasnapier.github.io/LEAVES/.

- **Software Name:** LEAVES
- **Year First Available:** 2024
- **Hardware Required:** Standard Computing Systems
- **Software Required:** Internet Browser (tested in Mozilla FireFox, Google Chrome and Microsoft Edge)
- **Availability and Cost:** Open Source, Free
- **Program Language:** Python3
- **Download Link:** https://thomasnapier.github.io/LEAVES/
- **Repository:** https://github.com/thomasnapier/LEAVES
- **Access Form:** Downloadable
- **License:** GNU GPL2

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix. Internal clustering performance

See Table A.4.

Data availability

Data will be made available on request.

References

Aide, T.M., Corrada-Bravo, C., Campos-Cerqueira, M., Milan, C., Vega, G., Alvarez, R., 2013. Real-time bioacoustics monitoring and automated species identification. PeerJ 1, e103.

Allaoui, M., Kherfi, M.L., Cheriet, A., 2020. Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. In: International Conference on Image and Signal Processing. Springer, pp. 317–325.

Allen-Ankins, S., McKnight, D.T., Nordberg, E.J., Hoefer, S., Roe, P., Watson, D.M., McDonald, P.G., Fuller, R.A., Schwarzkopf, L., 2023. Effectiveness of acoustic indices as indicators of vertebrate biodiversity. Ecol. Indic. 147, 109937.

Alonso, J.B., Cabrera, J., Shyamnani, R., Travieso, C.M., Bolaños, F., García, A., Villegas, A., Wainwright, M., 2017. Automatic anuran identification using noise removal and audio activity detection. Expert Syst. Appl. 72, 83–92.

Araya-Salas, M., Smith-Vidaurre, G., 2017. Warbler: an r package to streamline analysis of animal acoustic signals. Methods Ecol. Evol. 8 (2), 184–191.

Baltes, S., Ralph, P., 2020. Sampling in software engineering research: A critical review and guidelines.

Barber, J.R., Burdett, C.L., Reed, S.E., Warner, K.A., Formichella, C., Crooks, K.R., Theobald, D.M., Fristrup, K.M., 2011. Anthropogenic noise exposure in protected natural areas: estimating the scale of ecological consequences. Landsc. Ecol. 26, 1281–1295.

BatExplorer, 2020. BatExplorer (version 2.1)[computer program].

Boersma, P., Weenink, D., 2001. PRAAT, a system for doing phonetics by computer. Glot Int. 5, 341–345.

Brown, A., Garg, S., Montgomery, J., 2018. Scalable preprocessing of high volume environmental acoustic data for bioacoustic monitoring. PLoS ONE 13, e0201542.

Busleiman, A., Crook, J., Danneberg, R., Daulton, S., Kozikowski, R., Licameli, P., Sampson, P., Wharrie, B., 2020. Audacity 2.3.3.

Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. 3 (1), 1–27.

Cannam, C., Landone, C., Sandler, M., 2010. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In: Proceedings of the ACM Multimedia 2010 International Conference. Firenze, Italy, pp. 1467–1468.

Cartwright, M., Dove, G., Méndez Méndez, A.E., Bello, J.P., Nov, O., 2019. Crowdsourcing multi-label audio annotation tasks with citizen scientists. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. pp. 1–11.

Cartwright, M., Seals, A., Salamon, J., Williams, A., Mikloska, S., MacConnell, D., Law, E., Bello, J.P., Nov, O., 2017. Seeing sound: Investigating the effects of visualizations and complexity on crowdsourced audio annotations. Proc. ACM Hum.- Comput. Interact. 1 (CSCW), 1–21.

Charif, R., Waack, A., Strickman, L., 2010. Raven Pro 1.4 User's Manual. Vol. 25506974, Cornell Lab of Ornithology, Ithaca, NY.

Chase, J.M., Blowes, S.A., Knight, T.M., Gerstner, K., May, F., 2020. Ecosystem decay exacerbates biodiversity loss with habitat loss. Nature 584 (7820), 238–243.

Dalmaijer, E., Nord, C., Astle, D., 2022. Statistical power for cluster analysis. BMC Bioinformatics 23.

Darras, K., Pérez, N., M, Dilong, L., Hanf-Dressler, T., Markolf, M., Wanger, T., 2023. Ecosound-web: an open-source, online platform for ecoacoustics. F1000Research 9 (1224).

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (2), 224–227.

Dufourq, E., Batist, C., Foquet, R., Durbach, I., 2022. Passive acoustic monitoring of animal populations with transfer learning. Ecol. Informatics 70, 101688.

Dunn, J.C., 1974. Well-separated clusters and optimal fuzzy partitions. J. Cybern. 4 (1), 95–104.

Farina, A., Gage, S.H., 2017. Ecoacoustics: A new science. In: Ecoacoustics. John Wiley & Sons, Ltd, pp. 1–11, chapter 1.

Fowlkes, E.B., Mallows, C.L., 1983. A method for comparing two hierarchical clusterings. J. Amer. Statist. Assoc. 78 (383), 553–569.

Gan, H., Zhang, J., Towsey, M., Truskinger, A., Stark, D., van Rensburg, B.J., Li, Y., Roe, P., 2021. A novel frog chorusing recognition method with acoustic indices and machine learning. Future Gener. Comput. Syst. 125, 485–495.

Ghani, B., Denton, T., Kahl, S., Klinck, H., 2023. Global birdsong embeddings enable superior transfer learning for bioacoustic classification. Sci. Rep. 13 (1), 22876.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2018. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10 (2), 169–185.

Gonçalves-Souza, D., Verburg, P.H., Dobrovolski, R., 2020. Habitat loss, extinction predictability and conservation efforts in the terrestrial ecoregions. Biol. Cons. 246, 108579.

Grinfeder, E., et al., 2022. Soundscape dynamics of a cold protected forest: dominance of aircraft noise. Landsc. Ecol. 37 (2), 567–582.

Guerrero, M.J., Bedoya, C.L., López, J.D., Daza, J.M., Isaza, C., 2023. Acoustic animal identification using unsupervised learning. Methods Ecol. Evol. 14 (6), 1500–1514.

Hao, Z., Zhan, H., Zhang, C., Pei, N., Sun, B., He, J., Wu, R., Xu, X., Wang, C., 2022. Assessing the effect of human activities on biophony in urban forests using an automated acoustic scene classification model. Ecol. Indic. 144, 109437.

Harris, C.R., Millman, K.J., Van Der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., et al., 2020. Array programming with NumPy. Nature 585 (7825), 357–362.

Hoefer, S., et al., 2023. Passive acoustic monitoring in terrestrial vertebrates: a review. Bioacoustics 32 (5), 506–531.

Huang, X., Wu, L., Ye, Y., 2019. A review on dimensionality reduction techniques. Int. J. Pattern Recognit. Artif. Intell. 33 (10), 1950017.

Hubert, L., Arabie, P., 1985. Comparing partitions. J. Classification 2, 193–218.

Hunter, J.D., 2007. Matplotlib: A 2D graphics environment. Comput. Sci. Eng. 9 (3), 90–95.

Ishibashi, T., Nakao, Y., Sugano, Y., 2020. Investigating audio data visualization for interactive sound recognition. In: Proceedings of the 25th International Conference on Intelligent User Interfaces. pp. 67–77.

Joshi, K.A., Mulder, R.A., Rowe, K.M., 2017. Comparing manual and automated species recognition in the detection of four common south-east Australian forest birds from digital field recordings. Emu- Austral Ornithol. 117 (3), 233–246.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Informatics 61, 101236.

Kholghi, M., Phillips, Y., Towsey, M., Sitbon, L., Roe, P., 2018. Active learning for classifying long-duration audio recordings of the environment. Methods Ecol. Evol. 9 (9), 1948–1958.

Lasseck, M., 2019. Audio-based bird species identification with deep convolutional neural networks. In: Proceedings of the Working Notes of CLEF 2021. p. 11.

LeBien, J., Zhong, M., Campos-Cerqueira, M., Velev, J.P., Dodhia, R., Ferres, J.L., Aide, T.M., 2020. A pipeline for identification of bird and frog species in tropical soundscape recordings using a convolutional neural network. Ecol. Inform. 59, 101113.

Linke, S., Deretic, J.-A., 2019. Ecoacoustics can detect ecosystem responses to environmental water allocations. Freshwater Biol. 65, 133–141.

Linke, S., Gifford, T., Desjonquères, C., Tonolla, D., Aubin, T., Barclay, L., Karaconstantis, C., Kennard, M.J., Rybak, F., Sueur, J., 2018. Freshwater ecoacoustics as a tool for continuous ecosystem monitoring. Front. Ecol. Environ. 16 (4), 231–238.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., Wu, S., 2013. Understanding and enhancement of internal clustering validation measures. IEEE Trans. Cybern. 43 (3), 982–994.

Van der Maaten, L., Hinton, G., 2008. Visualizing data using t-SNE. J. Mach. Learn. Res. 9 (11).

Maćkiewicz, A., Ratajczak, W., 1993. Principal components analysis (PCA). Comput. Geosci. 19 (3), 303–342.

MacQueen, J., et al., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. Vol. 1, Oakland, CA, USA, pp. 281–297.

Marchal, J., Fabianek, F., Aubry, Y., 2022. Software performance for the automated identification of bird vocalisations: the case of two closely related species. Bioacoustics 31 (4), 397–413.

McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O., 2015. Librosa: Audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference. Vol. 8, pp. 18–25.

McInnes, L., Healy, J., Saul, N., Großberger, L., 2018. UMAP: Uniform manifold approximation and projection. J. Open Source Softw. 3 (29), 861.

McKinney, W., et al., 2010. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. Vol. 445, Austin, TX, pp. 51–56.

Mcloughlin, M.P., Stewart, R., McElligott, A.G., 2019. Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. J. R. Soc. Interface 16 (155), 20190225.

Merchant, N.D., Fristrup, K.M., Johnson, M.P., Tyack, P.L., Witt, M.J., Blondel, P., Parks, S.E., 2015. Measuring acoustic habitats. Methods Ecol. Evol. 6 (3), 257–265.

Mesaros, A., Heittola, T., Benetos, E., Foster, P., Lagrange, M., Virtanen, T., Plumbley, M.D., 2018. Detection and classification of acoustic scenes and events: Outcome of the DCASE 2016 challenge. IEEE/ACM Trans. Audio Speech Lang. Process. 26, 379–393.

Mosqueira-Rey, E., Hernández-Pereira, E., Alonso-Ríos, D., Bobes-Bascarán, J., Fernández-Leal, Á., 2023. Human-in-the-loop machine learning: a state of the art. Artif. Intell. Rev. 56 (4), 3005–3054.

Napier, T., Ahn, E., Allen-Ankins, S., Lee, I., 2024a. An optimised grid search based framework for robust large-scale natural soundscape classification. In: Liu, T., Webb, G., Yue, L., Wang, D. (Eds.), AI 2023: Advances in Artificial Intelligence. Springer Nature Singapore, Singapore, pp. 468–479.

Napier, T., Ahn, E., Allen-Ankins, S., Schwarzkopf, L., Lee, I., 2024b. Advancements in preprocessing, detection and classification techniques for ecoacoustic data: A comprehensive review for large-scale passive acoustic monitoring. Expert Syst. Appl. 252, 124220.

Nocera, T., Ford, W.M., Silvis, A., Dobony, C.A., 2019. Let's agree to disagree: Comparing auto-acoustic identification programs for northeastern bats. J. Fish Wildl. Manag. 10 (2), 346–361.

Parra-Hernández, R.M., Posada-Quintero, J.I., Acevedo-Charry, O., Posada-Quintero, H.F., 2020. Uniform manifold approximation and projection for clustering taxa through vocalizations in a neotropical passerine (rough-legged tyrannulet, Phyllomyias burmeisteri). Animals 10 (8), 1406.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Pettersson, L., 2004. Bat-specific signal analysis software: BatSound. In: Bat Echolocation Research: Tools, Techniques and Analysis. p. 130.

Phillips, Y.F., Towsey, M., Roe, P., 2018. Revealing the ecological content of long-duration audio-recordings of the environment through clustering and visualisation. PLoS ONE 13 (3), 1–27.

Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H., Pieretti, N., 2011. Soundscape ecology: The science of sound in the landscape. BioScience 61 (3), 203–216.

Powers, R.P., Jetz, W., 2019. Global habitat loss and extinction risk of terrestrial vertebrates under future land-use-change scenarios. Nat. Clim. Chang. 9 (4), 323–329.

Quinn, C.A., Burns, P., Gill, G., Baligar, S., Snyder, R.L., Salas, L., Goetz, S.J., Clark, M.L., 2022. Soundscape classification with convolutional neural networks reveals temporal and geographic patterns in ecoacoustic data. Ecol. Indic. 138, 108831.

Robert, J., Webbie, M., et al., 2018. Pydub.

Roch, M.A., Batchelor, H., Baumann-Pickering, S., Berchok, C.L., Cholewiak, D., Fujioka, E., Garland, E.C., Herbert, S., Hildebrand, J.A., Oleson, E.M., et al., 2016. Management of acoustic metadata for bioacoustics. Ecol. Inform. 31, 122–136.

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 Ecol. Evol. 12 (10), 1802–1808.

Ross, S.R.P.-J., O'Connell, D.P., Deichmann, J.L., Desjonquères, C., Gasc, A., Phillips, J.N., Sethi, S.S., Wood, C.M., Burivalova, Z., 2023. Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions. Funct. Ecol. 37 (4), 959–975.

Rousseeuw, P.J., 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65.

Salamon, J., Bello, J.P., Farnsworth, A., Robbins, M., Keen, S., Klinck, H., Kelling, S., 2016. Towards the automatic classification of avian flight calls for bioacoustic monitoring. PLoS ONE 11, e0166866.

Schubert, E., Sander, J., Ester, M., Kriegel, H.P., Xu, X., 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. ACM Trans. Database Syst. 42 (3), 1–21.

Shammamah, H., 2019. Visualization of bioinformatics data with dash bio. In: Proceedings of the 18th Python in Science Conference. pp. 126–133.

Shinners, P., 2011. Pygame.

Sjölander, K., Beskow, J., 2000. Wavesurfer - an open source speech tool. In: Proc. 6th International Conference on Spoken Language Processing. ICSLP 2000, Vol. 4, pp. 464–467.

Solsona-Berga, A., Frasier, K.E., Baumann-Pickering, S., Wiggins, S.M., Hildebrand, J.A., 2020. DetEdit: A graphical user interface for annotating and editing events detected in long-term acoustic monitoring data. PLoS Comput. Biol. 16 (1), e1007598.

Specht, R., 2002. Avisoft-saslab pro: sound analysis and synthesis laboratory. Avis. Bioacoustics Berl. 2002, 1–723.

Stowell, D., 2022. Computational bioacoustics with deep learning: a review and roadmap. PeerJ 10, e13152.

Stowell, D., Sueur, J., 2020. Ecoacoustics: acoustic sensing for biodiversity monitoring at scale. Remote. Sens. Ecol. Conserv. 6 (3), 217–219.

Strehl, A., Ghosh, J., 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3 (Dec), 583–617.

Sueur, J., Aubin, T., Simonis, C., 2008. Seewave, a free modular tool for sound analysis and synthesis. Bioacoustics 18 (2), 213–226.

Sueur, J., Farina, A., 2015. Ecoacoustics: the ecological investigation and interpretation of environmental sound. Biosemiotics 8 (3), 493–502.

Szewczak, J., 2010. SonoBat (version 3)[computer program].

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., Roe, P., 2014. Visualization of long-duration acoustic recordings of the environment. Procedia Comput. Sci. 29, 703–712.

Truskinger, A., Cottman-Fields, M., Eichinski, P., Towsey, M., Roe, P., 2014. Practical analysis of big acoustic sensor data for environmental monitoring. In: 2014 IEEE Fourth International Conference on Big Data and Cloud Computing. pp. 91–98.

Van Parijs, S.M., et al., 2015. NEPAN: A U.S. northeast passive acoustic sensing network for monitoring, reducing threats and the conservation of marine animals. Mar. Technol. Soc. J. 49 (2), 70–86.

Vella, K., Capel, T., Gonzalez, A., Truskinger, A., Fuller, S., Roe, P., 2022. Key issues for realizing open ecoacoustic monitoring in Australia. Front. Ecol. Evol. 9.
Villanueva-Rivera, L.J., Pijanowski, B.C., 2012. Pumilio: a web-based management system for ecological recordings. Bull. Ecol. Soc. Am. 93 (1), 71–81.
Wang, M., Mei, J., Darras, K.F., Liu, F., 2023. Vggish-based detection of biological sound components and their spatio-temporal variations in a subtropical forest in eastern China. PeerJ 11, e16462.
WildlifeAcoustics, 2020. Kaleidoscope pro 5 user guide.
Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., et al., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci. Data 3 (1), 1–9.
Wood, C.M., Kahl, S., 2024. Guidelines for appropriate use of BirdNET scores and other detector outputs. J. Ornithol. 1–6.