Interactive Machine Learning Solutions for Acoustic Monitoring of Animal Wildlife in Biosphere Reserves

Thiago S. Gouvêa[1,2] , **Hannes Kath**[1,2] , **Ilira Troshani**[1,2] , **Bengt Lüers**[1] , **Patrícia P. Serafini**[3] , **Ivan B. Campos**[3,4] , **André S. Afonso**[5] , **Sérgio M. F. M. Leandro**[6] , **Lourens Swanepoel**[7] , **Nicholas Theron**[8] , **Anthony M. Swemmer**[9] and **Daniel Sonntag**[1,2]

[1]German Research Center for Artificial Intelligence (DFKI), Oldenburg, Germany
[2]Applied Artificial Intelligence, University of Oldenburg, Germany
[3]National Center for Wild Bird Conservation and Research (CEMAVE), Chico Mendes Institute for Biodiversity Conservation (ICMBio), Brazil
[4]Departamento de Biologia Geral, Universidade Federal de Minas Gerais (UFMG), Brazil
[5]Marine and Environmental Research Centre (MARE), Aquatic Research Network, Department of Life Sciences, University of Coimbra, Portugal
[6]School of Tourism and Maritime Technology, Polytechnic of Leiria, Peniche, Portugal
[7]Department of Biological Sciences, Faculty of Science, Engineering and Agriculture, University of Venda, South Africa
[8]Kruger to Canyons Biosphere Region NPC, South Africa
[9]South African Environmental Observation Network (SAEON), Kruger National Park, South Africa
{thiago.gouvea, hannes_berthold.kath, ilira.troshani, bengt.lueers, daniel.sonntag}@dfki.de
{patricia.serafini, ivan.campos}@icmbio.gov.br
afonso.andre@gmail.com
sergio.leandro@ipleiria.pt
lourens.swanepoel@univen.ac.za
nicktheron@kruger2canyons.org
am.swemmer@saeon.nrf.ac.za

Abstract

Biodiversity loss is taking place at accelerated rates globally, and a business-as-usual trajectory will lead to missing internationally established conservation goals. Biosphere reserves are sites designed to be of global significance in terms of both the biodiversity within them and their potential for sustainable development, and are therefore ideal places for the development of local solutions to global challenges. While the protection of biodiversity is a primary goal of biosphere reserves, adequate information on the state and trends of biodiversity remains a critical gap for adaptive management in biosphere reserves. Passive acoustic monitoring (PAM) is an increasingly popular method for continued, reproducible, scalable, and cost-effective monitoring of animal wildlife. PAM adoption is on the rise, but its data management and analysis requirements pose a barrier for adoption for most agencies tasked with monitoring biodiversity. As an interdisciplinary team of machine learning scientists and ecologists experienced with PAM and working at biosphere reserves in marine and terrestrial ecosystems on three different continents, we report on the co-development of interactive machine learning tools for semi-automated assessment of animal wildlife.

1 Introduction

Biodiversity loss is increasing rapidly, and the importance of its conservation is expressed in the United Nations (UN) Sustainable Development Goals (SDGs) 14 (Life Below Water) and 15 (Life On Land). The UN Conference of the Parties to the Convention on Biological Diversity (CBD) has set a biodiversity conservation framework with goals and targets for the current decade [CBD, 2022], and experts estimate that meeting those targets will require transformative change [IPBES, 2019].

Established by UNESCO's Man and the Biosphere programme, biosphere reserves are ideal sites for testing new approaches to understanding and managing biodiversity [UNESCO, 2019]. As biosphere reserves include terrestrial and marine ecosystems, they allow for the development of local solutions to global challenges. Biosphere reserves are composed of core protected areas surrounded by zones of various degrees of human development, and most have established networks linking researchers and land managers (both inside and outside of protected areas). The core protected areas provide habitats that buffer many species against extinction while also providing essential ecosystem services to humans [Watson *et al.*, 2014]. As indicated by the goal set by CBD of protecting 30% of all terrestrial, inland water, marine and coastal ecosystems [CBD, 2022], protected areas are globally recognised as essential for conservation. The surrounding zones of a biosphere reserve are meant for economic activities based on sound ecological practices, fostering socio-culturally and ecologically sustainable economic development [UNESCO, 2019].

While the protection of biodiversity is a primary goal of biosphere reserves, adequate information on the state and trends of biodiversity remains a critical gap for adaptive management of biosphere reserves and global biodiversity assessments. Adaptive management is a learning-based approach to planning and managing natural resource systems. It is based on a perception that managed ecosystems are complex, dynamic systems with a large degree of unpredictability [Gunderson, 2008]. As such, they generate compelling AI challenges, and research in interactive machine learning can make important contributions to help address them. However, to harness the potential of this research, it is crucial to establish proper data generation and management processes.

Biodiversity is a broad concept, encompassing multiple dimensions of diversity (from genes to ecosystems) and is both conceptually difficult and expensive to measure and monitor. In the present proposal, we define the scope at the species and ecosystem level monitoring of animal wildlife. Traditional methods of animal wildlife monitoring rely on specialists being physically present on the monitored locations (e.g., [Plumptre, 2000]). More recently, passive acoustic monitoring (PAM)—the use of audio recording devices to capture sounds and vocalisations in the environment in a minimally invasive manner—has emerged as an alternative for continuous, reproducible, scalable, and cost-effective monitoring of animal wildlife [Sugai *et al.*, 2019]. While PAM has been widely used for research over the past decade, it has yet to be adopted by managers of protected areas or agencies tasked with long-term monitoring of biodiversity at national or regional scales. The amount of data produced by PAM is large enough to demand automation, raising data management and analysis requirements that pose a barrier for adoption.

As an interdisciplinary team of machine learning scientists and ecologists experienced with PAM and working at biosphere reserves in marine and terrestrial ecosystems on three different continents, we are co-developing interactive machine learning tools for analysis of PAM datasets. In section 2, we provide a brief overview of the existing tools for PAM data analysis, and introduce the methods being co-developed by our teams. In section 3 we describe the context in which these tools will be deployed to support adaptive management of biosphere reserves. In section 4, we discuss the impact and relevance of our project in light of the UN SDGs.

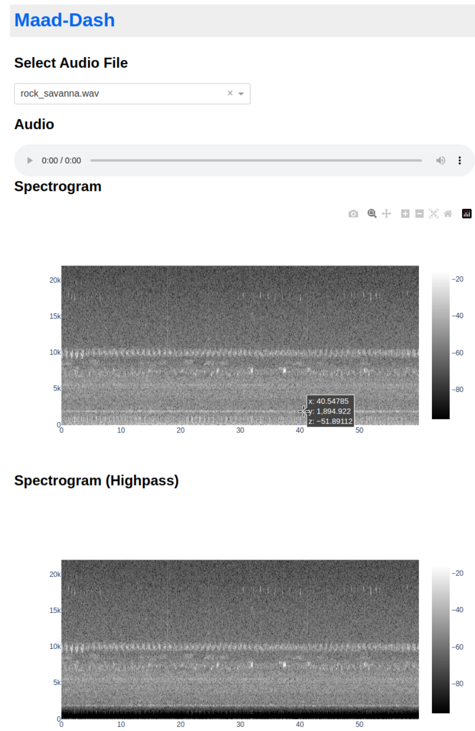


Figure 1: **Proof-of-concept interactive interface for scikit-maad**. The prototype will include other important functionalities from the scikit-maad library such as spectrogram segmentation and event clustering. (Available at the time of publication at https://gitlab-1137-main-3r37sitkfa-lz.a.run.app/)

2 Methods

The emergence of PAM raises new challenges to the ecology community, and to biosphere reserve management in particular. Information extraction is often done manually with software such as Audacity[1] or Label-Studio[2]: domain experts listen to each audio file, annotating events by selecting time segments on a graphical representation of the sound (e.g., amplitude envelope or spectrogram) with point-click-drag actions. This approach is laborious and incompatible with continued, wide coverage monitoring due to the large volume of data generated by PAM.

Several solutions for efficient analysis of PAM datasets have been proposed. DetEdit is a machine learning-free tool that allows simultaneous detection of bouts of events through a configurable signal processing pipeline that includes a GUI for accepting/rejecting detections [Solsona-Berga *et al.*, 2020]; it runs on a proprietary platform, and has only been evaluated on odontocete echolocation click datasets. scikit-maad is a tool for large scale PAM data analysis by spectrogram segmentation and clustering [Ulloa *et al.*, 2021]; as a command line tool, it lacks interactivity (but see section 2.1). Another set of solutions consists of fully automated detection software (e.g., [Gillespie *et al.*, 2009; Heinicke *et al.*, 2015; Kahl *et al.*, 2021]). A related alternative consists of train-

[1] https://www.audacityteam.org/
[2] https://labelstud.io/

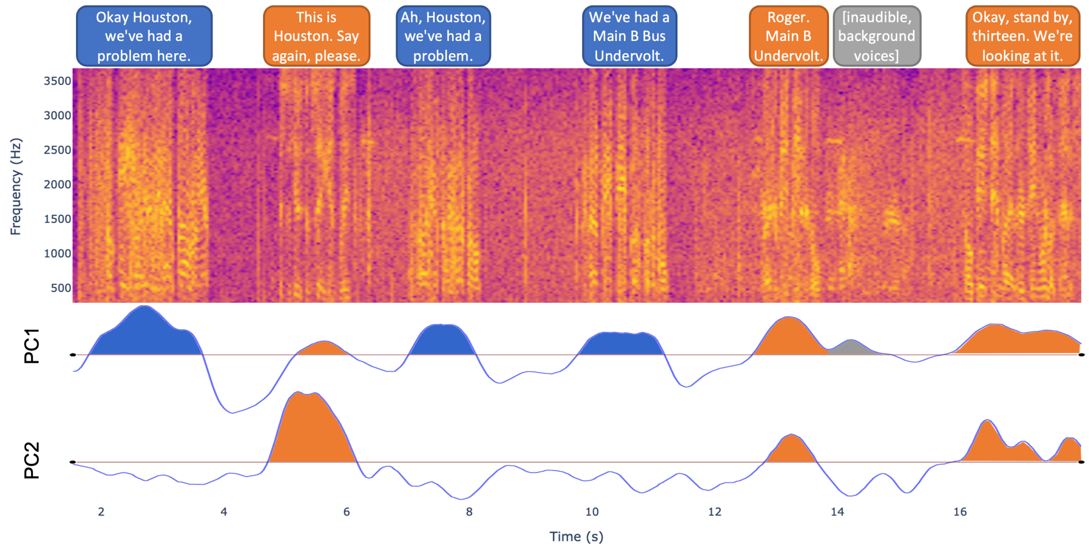


Figure 2: **A toy example of graphical data programming.** Visual inspection of latent dimensions of a linear autoencoder reveals interpretable features. *(top)* A dialog between astronauts aboard Apollo 13 and Houston base. Colours distinguish between speakers. *(middle)* Spectrogram of the dialog recording. *(bottom)* First two principal components of the spectrogram (preceded by temporal smoothing). Hand-set thresholds isolate event occurrences (PC1: voice vs background; PC2: selective for Houston voices). Data source: https://en.wikipedia.org/wiki/File:Apollo13-wehaveaproblem.ogg.

ing custom supervised machine learning methods, including deep learning models (e.g., [Stowell, 2022]). This approach requires expertise in designing the models, and is often limited by the lack of annotated datasets for training. Fully automated solutions are often non-interpretable and community-specific, with geographically restricted demonstrated validity (often tailored to ecosystems located in the global North.) Furthermore, while automated detection can in principle deliver high productivity gains, these gains are often offset by the independent verification efforts required to develop trust in them, as well as performing the adjustments they require when applied to previously unseen datasets (e.g., in a geographical location not represented in the training set). For these reasons among others, experts who are responsible for important decisions are often hesitant to rely solely on hands-off methods [Rudin, 2019], and intermediate levels of automation may be more desirable [Parasuraman *et al.*, 2000; Van Zoelen *et al.*, 2023].

Interactivity can help alleviate these problems by shortening the cycle of verification (developing trust) and fine tuning (feeding expert knowledge back into the model) [Amershi *et al.*, 2014; Tusfiqur *et al.*, 2022]. While acknowledging the importance of machine learning research driven by developing novel algorithms evaluated based on their accuracy on annotated benchmark datasets, the research field of interactive machine learning emphasises improving the effectiveness of machine learning tools that are designed and utilised by domain experts in their areas of expertise, where qualities such as interpretability may be more crucial [Simard *et al.*, 2017; Hartmann *et al.*, 2022].

In the present project we set out to co-develop a set of interactive machine learning tools for facilitating analysis of PAM datasets by designing intelligent user interfaces (IUIs) that integrate interpretable machine learning models with interactive interfaces [Zacharias *et al.*, 2018]. We start by proposing a template for raising the impact of existing tools by making them interactive (section 2.1). Next, we propose methods for data annotation by hand-designing interpretable features (section 2.2), or by harnessing explainable methods to extract strong labels from existing weakly annotated collections (section 2.3). Finally, we propose forms of interacting with deep neural networks in their own inner space (section 2.4).

2.1 Serving Existing Tools Interactively

While the acquisition, handling and storing of audio data can be very cost-effective, relying on human labor to annotating audio data is very inefficient. Some tasks of the data annotation process can be automated, which relieves the workload of human annotators. For these use cases, the Python package scikit-maad has been proposed, which implements functionality to automate audio data annotation [Ulloa *et al.*, 2021]. scikit-maad requires its users to write applications in the Python programming language, which create static plots of the audio data annotation steps. scikit-maad aims to provide a low-code environment for these use cases. To enable users without any programming knowledge to participate in the audio annotation workflow, we will implement a no-code interactive user interface for automation-assisted audio annotation based on scikit-maad [Ulloa *et al.*, 2021] and Plotly Dash[3]. Our proof-of-concept of such a scikit-maad application replaces the static matplotlib plots with Plotly plots, handles their integration using React.js and serves an interactive web application using flask, as can be seen in figure 1. This web service has been containerised as a Docker image and deployed using Google Cloud Run. Note that by using web

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

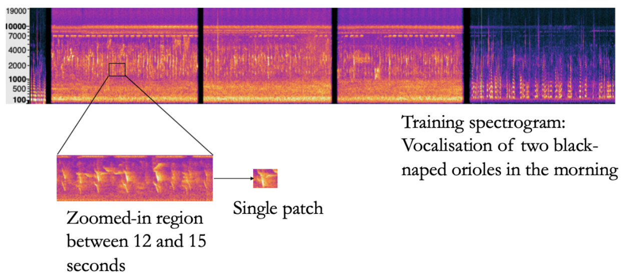
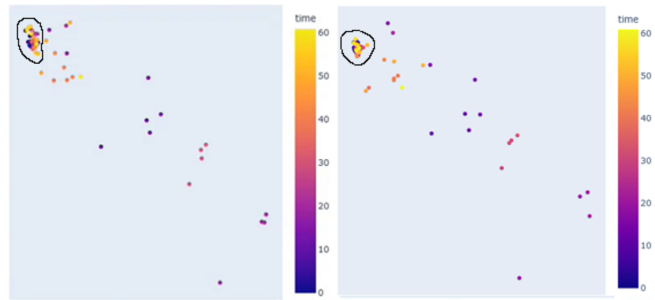


Figure 3: **Spectrogram representation of a sample from an animal sound collection.** The target species is black-naped oriole bird (*Oriolus chinensis*). The first 5 and the last 30 seconds of the recordings consist of human speech, while many different co-occurring species are heard in the background. The zoomed-in region highlights the target vocalisations, and the highlighted patch illustrates a distinctive feature learned by local prototype models. (Recording kindly provided by Fonoteca Neotropical Jacques Vielliard).



(a) Before re-training (b) After re-training

Figure 4: **Actionable visualisation of PAM data in latent space of a deep neural network.** State-space representation of the fine-tuned model before and after annotating one file and re-training the deep generative model (50 epochs). The gain for annotation is visible in the upper left corner, where retraining increases the distance between non-events (circled group of points) and events (outliers), making it easier for the user to distinguish between points that are non-events and points that are events.

technologies, the delivery process from our web server in the cloud to the web browser of the user is as trivial as opening a link and loading the website. Once they loaded the website, the users of our proof-of-concept application can select an audio file from a list of predefined files. The system will then plot a simple spectrogram of the audio data contained in the selected file. The user can then zoom and pan the spectrogram interactively using plotly's interaction mechanisms.

The tool mentioned above exemplifies a broader objective within the proposed research project: to promote evidence-based biodiversity management by delivering state-of-the-art machine learning solutions to domain experts in a timely and user-centric manner in the form of interactive web applications.

2.2 Graphical Data Programming

Data programming is a data annotation approach based on designing a set of heuristic labelling functions, each of which may have low accuracy but wide coverage. Compared to manual annotation of individual event occurrences, this approach promises higher efficiency [Ratner *et al.*, 2016].

We have proposed a graphical implementation of data programming for PAM data based on exposing feature design primitives on a graphical user interface (GUI) [Gouvêa *et al.*, 2022a; Gouvêa *et al.*, 2022b]. Experimenting with these primitives is facilitated by incremental and reversible actions that provide immediate feedback. When an informative feature is identified (i.e. one whose activation is indicative of occurrence of an event of interest), data can be efficiently annotated by setting a threshold on that feature's activation level. Figure 2 illustrates this principle using a toy model and dataset.

2.3 Semi-Supervised Learning With Local Prototype Networks

Existing animal sound collections such as UNICAMP's Fonoteca Neotropical Jacques Vielliard (FNJV)[4] or Cornell University's Macaulay Library[5] constitute valuable resources. However, leveraging these collections to provide the annotated data for training supervised machine learning algorithms poses technical challenges as these datasets are often weakly annotated (i.e., they are labelled at the file level, rather than identifying time segments containing the calls of the species of interest); furthermore, they often contain multiple signals alongside the target one, such as human speech and other co-occurring species (figure 3).

To address this challenge, we plan to extract the relevant signals by repurposing explainable machine learning methods based on learning local patterns (prototypes) that are representative of the classes of interest [Biehl *et al.*, 2016; Chen *et al.*, 2019; Brendel and Bethge, 2019]. For an initial evaluation, we trained a compact model inspired on ProtoPNet [Chen *et al.*, 2019] on the MNIST dataset, achieving a validation accuracy of 95.9 %. The next step will be to preprocess the data from the FNJV collection by transforming them into spectrograms, feed them to our neural prototype model and visualise the learned prototypes that lead to the predictions. In line with our human-in-the-loop approach, an interactive interface will facilitate triage of learned prototypes by a human expert.

By learning local patches, either in input- [Brendel and Bethge, 2019] or latent space [Chen *et al.*, 2019], we expect to be able to extract the species-identifying calls from the multi-signal files that make up the collections, generating strong (timestamp level) annotations from weakly (file level) annotated data, thus rendering these datasets amenable to supervised learning methods.

2.4 Interactive Representation Learning

In all tools listed above, the user interacts with the system in input space, i.e., by operating on visual representations of the sound. However, interaction with deep learning models

[4]https://www2.ib.unicamp.br/fnjv/

[5]https://www.macaulaylibrary.org

is also possible in latent space, at the level where the 'reasoning' of the model (its internal representations, transformations, and learning process) takes place. We are developing tools that explore this potential through interactive representation learning.

Representation learning refers to a set of machine learning techniques to automatically discover and construct meaningful features from raw, high-dimensional data. The objective is to learn a compressed and abstract representation of the data that captures the underlying structure and relationships present in the input [Bengio *et al.*, 2013]. There are different approaches to representation learning, including unsupervised, semi-supervised, and supervised methods. Unsupervised techniques learn representations without explicit labels or supervision (e.g., variational autoencoders (VAEs) [Kingma and Welling, 2013]). Semi-supervised methods leverage a small amount of labelled data along with unlabelled data to improve the learned representations (e.g., [Kingma *et al.*, 2014; Siddharth *et al.*, 2017]). Supervised methods, on the other hand, utilise labelled data to learn representations that are directly optimised for a specific task. In the situations represented in this proposal, where annotations are scarce or unavailable, unsupervised and semi-supervised methods become particularly relevant and applicable.

Our goal is to investigate effective strategies for utilising these techniques to generate annotations through interactive graphics. We have previously used compact learned representations to facilitate data annotation with an IUI. In [Prange and Sonntag, 2021], activations of the hidden layers of a convolutional neural network (CNN) trained on an image classification task were projected down to a 3D space with unsupervised methods. This representation was then rendered in virtual reality where the user could use hand gestures to annotate the data. The annotations were used to fine tune the CNN, leading to a new 3D visualisation in an iterative workflow. In the current project, we propose to extend that tool by leveraging semi-supervised methods to jointly learn a compact representation that captures both the structure of the input data and the user-added annotations [Kath *et al.*, 2023a; Kath *et al.*, 2023a; Kath *et al.*, 2023b; Kath *et al.*, 2023c].

In [Kath *et al.*, 2023b], we present a tool that offers three ways to annotate PAM audio files. First, it allows experts to draw bounding boxes directly on the spectrogram, as is common practice in the field. Second, the user can request suggestions for bounding boxes from a semi-supervised deep generative model fit to the data [Kath *et al.*, 2023a]. Lastly, the user can interact directly with the latent space of the semi-supervised model by lasso-selecting one or more points in the state-space representation, where each point represents one short time segment of the audio file. This creates a corresponding table of all selected times around which the user can create bounding boxes using a button. After creating new annotations, the user can re-train the deep generative model to increase the prediction accuracy of the proposed bounding boxes and the margin between events and non-events in the state-space representation, as shown in figure 4. We conducted preliminary evaluations of bounding box prediction and usability of state-space representations using a synthetic dataset, resulting in a fine-tuned model with a bounding box prediction accuracy of 79.9 % and an F-score of 94.2 % (for more information refer to [Kath *et al.*, 2023b]).

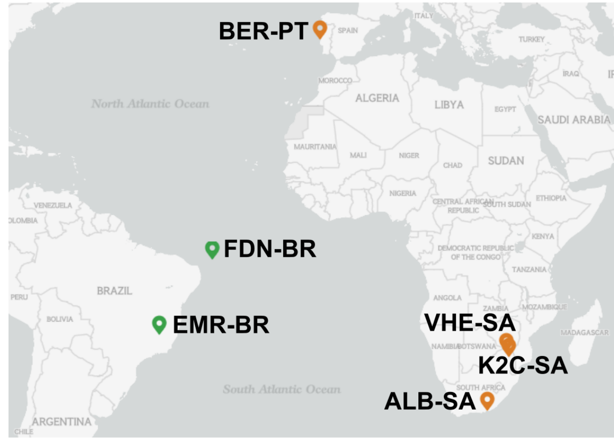


Figure 5: **Location of the biosphere reserves and protected areas**. Data acquisition is either ongoing (green) or planned (orange). (Credit: NatGeo Mapmaker)

3 Implementation Plan

In the previous section, we described the interactive machine learning tools that are being developed for efficient annotation of PAM datasets. Next, we describe the context in which these tools will be deployed, and how they will support adaptive management in biosphere reserves. We also provide a timeline of the proposed activities and expected outputs in figure 6.

3.1 Data Acquisition

We aim to establish demonstration sites for PAM-based monitoring of biodiversity in both protected areas and adjacent land uses, in marine and terrestrial ecosystems in 3 biosphere reserves on 3 continents (figure 5).

Active PAM data acquisition is currently ongoing at Espinhaço Mountain Range, Brazil (EMR-BR) and Fernando de Noronha, Brazil (FDN-BR)[6]. Drawing from this experience and following engagement with relevant managers of protected areas and landowners at different sites, we aim to establish multiyear PAM programs in Berlengas Biosphere Reserve, Portugal (BER-PT), Kruger-to-Canyons Biosphere Reserve, South Africa (K2C-SA), Vhembe Biosphere Reserve, South Africa (VHE-SA), and Algoa Bay, South Africa (ALB-SA). Acoustic sampling will be carried in terrestrial ecosystems at EMR-BR, K2C-SA, and VHE-SA, while marine recorders will be set up in the coastal areas of BER-PT and ALB-SA. Long-term research projects are already in place at all these sites, with direct involvement of at least one team member on this proposal. Networks to landowners and managers, and permission to access the sites are well established.

[6]Fernando de Noronha, Brazil (FDN-BR) is a designated protected area recognised by UNESCO as a World Natural Heritage site.

The data from these acoustic recorders will be used to develop optimal sampling designs for long-term monitoring of biodiversity in a variety of ecosystems. In addition, all data produced will be processed, archived, and disseminated in accordance with the FAIR principles [Wilkinson *et al.*, 2016] (see data management plan on section 3.5).

3.2 Tool Development

The overall goal for the tools described in section 2 is to extract variables of interest, as defined in the ecology lexicon, from raw acoustic data. For that purpose, tool development is being carried in line with principles of co-development [Woodall *et al.*, 2021], with domain experts as an integral part of the AI project and getting involved from an early stage in development.

Our team holds regular meetings that help guide initial design choices, and frequent feedback is decisive for prioritising tools, and adjusting for the needs of the domain expert in charge of biosphere reserve management. Of relevance for tool development is the fact that the concept of biodiversity encompasses multiple dimensions (from genes to ecosystems). Even after setting our scope to animal wildlife suitable for acoustic monitoring, observations of interest still span a range of abstraction levels—from syllables, to calls, to species, to communities, to soundscapes—thus posing challenges to the task of mapping from acoustic events to ecological phenomena. The latter, however, have been helpfully organised by the ecology community into a taxonomy of essential biodiversity variables (EBVs) [Pereira *et al.*, 2013; Jetz *et al.*, 2019]. By working together, we have identified the need for tools that are agnostic about the EBV level of abstraction: they should allow ecologists and biosphere reserve managers to set the level of granularity as required by the specific questions arising in the context of their adaptive management practice (see section 3.3 for a concrete example).

The tools described in section 2 are currently at early stages of development. We expect to have functional prototypes of all of them within the next months, and to have concluded the first user studies around first quarter of 2024 (see project timeline in figure 6). All software and source code will be made available on DFKI's public GitHub page[7].

3.3 Data Analysis

Adaptive management of biosphere reserves comprises the practice of performing integrative assessments and developing hypotheses to be evaluated through management actions. A necessary early step is to establish standardised methods for monitoring over large temporal and spatial scales.

Once PAM data acquisition is established in a given studied biosphere reserve, the tools described in section 2 will be used to characterise trends and dynamics of animal biodiversity, as well as to assess the impact of management decisions in an evidence-based manner.

As a concrete example, a new pest control initiative was carried at one of the sites that are part of this project. PAM data has been acquired before and after the eradication of the invasive species. Our tools will help assessing the effectiveness and impact of the program (e.g., change in composition of the communities). Another example is the lack of data in face of the occurrence of unpredictable disturbances such as wildfires. Although it is known that altered fire regimes have an important impact on biodiversity, managers of protected areas rarely have access to standardised data from before and after a wildfire to properly assess its impacts. A continuous monitoring program will provide data on the impacts of this unpredictable event, as well as other disturbances of natural and anthropogenic causes. Other specific questions will arise at each recording site as the data comes in—each with its own focus on, e.g., occurrence and abundance of species of conservation concern, or impact of different management actions. We expect to report on specific questions in the near future. Importantly, these reports will at first be in the form of joint studies involving the machine learning and domain (ecologists and biosphere reserve managers) expert partners. In the long run, we aim at domain experts developing full autonomy in using the tools to carry out the analyses required to address adaptive management questions that arise in their practice (figure 6).

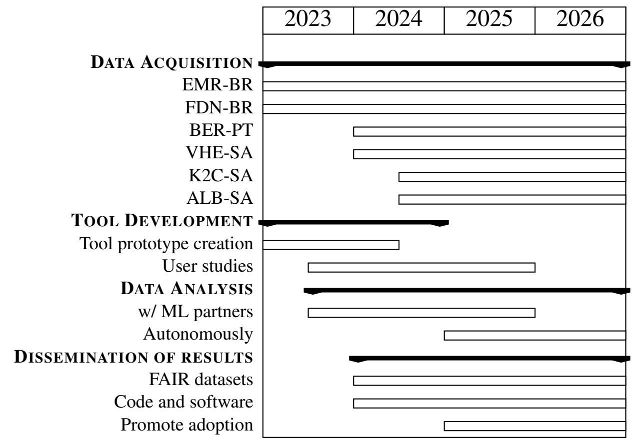


Figure 6: **Project timeline.**

3.4 Dissemination of Results

Besides making software, source code, and data publicly available, results from our work will be disseminated in other significant ways. We will engage with managers of protected areas and national and international agencies mandated to monitor biodiversity, at both the establishment and output phases of the project, to better understand barriers to the use of PAM-based biodiversity monitoring, and promote transnational adoption.

The Brazilian national biodiversity conservation agency (ICMBio) has implemented MONITORA, one of the largest biodiversity monitoring programmes in the world. With the main goal of measuring the effectiveness of protected areas to conserve biodiversity, the programme executes data sampling in various, extremely diverse ecosystems. Within the MONITORA program, the acoustic monitoring efforts are being led by one of our team members. The results of our project are

[7]https://github.com/DFKI-Interactive-Machine-Learning

going to inform decisions by ICMBio about extending PAM into some of the most diverse protected areas on the planet. Similar dissemination will take place in the other countries involved.

In South Africa, we will engage with the following stakeholders that are relevant in this regard: 1) The South African National Biodiversity Institute (SANBI), a government agency mandated to monitor and study changes in biodiversity at a national level); 2) the Oceans and Coast division of the national government Department of Forestry, Fisheries and Environment; 3) the Scientific Services department of SANParks (the South African authority responsible for managing all the national parks of South Africa); 4) the scientific division of MTPA (the state / provincial agency responsible for management of the BRCNR reserve. To encourage the use of the PAM protocol at a broader African scale, outside of South Africa, the Head of Science Support for Africa Parks will also be included in all engagements (Africa Parks is an NGO responsible for managing 22 national parks throughout Africa).

In Portugal, BER-PT already has a co-management plan running between private stakeholders (e.g., from fishing and tourism industries), academia, NGOs (e.g., birdwatchers society), and public authorities, and we will contribute to raising awareness about the ecological dynamics and resource management strategies in the biosphere reserve. The Polytechnic Institute of Leiria (IPL), represented in our team, has good ties with the public administration. The association of IPL with Smart Ocean Peniche (SOP), a business incubator, will provide a fruitful ecosystem for potential commercial routes for disseminating and scaling the methods developed in this project.

3.5 Data Management Plan

The proposed research project poses a series of challenges to the hardware and software infrastructure. Because the participants are distributed across multiple institutions, time zones and continents, close collaboration at a distance is paramount. Because the project participants are experts in a diverse set of domains, interdisciplinary work needs to be fostered. We will address these challenges by using tools and platforms that are universal and agnostic on every level. For example, our data set will be stored on Google Cloud storage, making it accessible globally without requiring any organization account. In addition, datasets will be indexed and published through the Global Biodiversity Information Facility (GBIF) in line with FAIR principles [Wilkinson *et al.*, 2016]. Source code and accompanying configuration will be stored on public repository, which is also a neutral third party to all the participants. This will avoid the development of information silos, as all participants get access on equal terms. Most of our software projects will be written in Python, as it has become the de-facto standard language of data science. This will lower the barrier of entry for collaboration on code artefacts. We will test our software projects using tox, which ensures setting up test environments repeatably. This will enable every researcher to run and debug any software of the whole research project locally. The accompanying documentation will be hosted on the third-party hosting service readthedocs.org. This will ensure maximum accessibility of any information necessary for working with the code and dataset artefacts. We will aim to also document our code by means of code itself and to that end include the examples to our code base in our test suites. Example applications are by the virtue of their graphicness more readable to novice users of a software package than the more abstract parts of the test suite.

4 Discussion

As a team of machine learning scientists, ecologists, and managers experienced with PAM and working at marine and terrestrial ecosystems in biosphere reserves in three different continents, we are co-developing interactive machine learning tools for PAM with the purpose to facilitate adaptive, evidence-based management of biosphere reserves and protected areas.

Protected areas are globally recognised as essential for conservation [CBD, 2022], and biodiversity management requires enhanced knowledge management. The widespread adoption of standardised PAM-based biodiversity monitoring in different ecosystem types would be a significant step forward for evidence-based management of biodiversity both inside and outside protected areas, with positive implications for SDGs 14 (life below water) and 15 (life on land).

Traditionally, each ecosystem type (forest, grassland, desert, freshwater, marine, etc.) has its own sampling method to generate information on biodiversity states and trends. To the best of our knowledge, there are no global initiatives on biodiversity monitoring that cover ecosystems as diverse as the ones covered in our project—namely forests, grasslands and marine ecosystems. In addition, the proposed transnational co-development of PAM-based biodiversity monitoring protocols and analysis tools is well in line with SDG 17 (Partnerships for the goals) and Target 20 of the Kunming-Montreal Biodiversity Framework, namely to "[s]trengthen capacity-building and development, access to and transfer of technology, [. . .] including through South-South, North-South and triangular cooperation, to meet the needs for effective implementation [. . .] commensurate with the ambition of the goals and targets of the framework." [CBD, 2022].

We are exploring different routes for designing interactive machine learning tools based on interpretable models—from a graphical user interface for design of interpretable features, to learning of interpretable prototypes, to actionable visualisations of the latent space of semi-supervised deep generative models—and we are developing a deployable, cloud-compatible data management and analysis infrastructure to serve tools—both developed in house and published by others—as user facing web applications. In contrast with the common machine learning practice of extracting expert knowledge by generating labels for model training with the goal of fully automating inference, our methods keep the user/domain expert in the loop to inspect and validate every sample while optimising for efficiency. This choice for a lower level of automation [Parasuraman *et al.*, 2000] seeks to implement a pattern of collaboration between domain experts and AI that harnesses the power of machine learning while privileging human judgment [Van Zoelen *et al.*, 2023].

Acknowledgments

This work has partly been supported by the Lower Saxony Ministry of Science and Culture, the Computational Sustainability and Technology research group (cst.dfki.de), the DFKI research department of interactive machine learning, and the Endowed Chair of Applied Artifical Intelligence (AAI) of the University of Oldenburg. This study was also financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil* (CAPES) - Finance Code 001, scholarship number 88887.695311/2022-00.

References

[Amershi *et al.*, 2014] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the People: The Role of Humans in Interactive Machine Learning. *AI Magazine*, 35(4):105–120, December 2014. Number: 4.

[Bengio *et al.*, 2013] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, August 2013. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[Biehl *et al.*, 2016] Michael Biehl, Barbara Hammer, and Thomas Villmann. Prototype-based models in machine learning. *WIREs Cognitive Science*, 7(2):92–111, 2016. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/wcs.1378.

[Brendel and Bethge, 2019] Wieland Brendel and Matthias Bethge. Approximating cnns with bag-of-local-features models works surprisingly well on imagenet. *arXiv:1904.00760*, 2019.

[CBD, 2022] CBD. Kunming-Montreal Global Biodiversity Framework, December 2022.

[Chen *et al.*, 2019] Chaofan Chen, Oscar Li, Daniel Tao, Alina Barnett, Cynthia Rudin, and Jonathan K Su. This Looks Like That: Deep Learning for Interpretable Image Recognition. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

[Gillespie *et al.*, 2009] Douglas Gillespie, David K. Mellinger, Jonathan Gordon, David McLaren, Paul Redmond, Ronald McHugh, Philip Trinder, Xiao-Yan Deng, and Aaron Thode. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *The Journal of the Acoustical Society of America*, 125(4):2547–2547, April 2009.

[Gouvêa *et al.*, 2022a] Thiago S Gouvêa, Ilira Troshani, Marc Herrlich, and Daniel Sonntag. Annotating Sound Events Through Interactive Design of Interpretable Features. *Proceedings of the 1st International Conference on Hybrid Human-Machine Intelligence*, pages 305–306, 2022.

[Gouvêa *et al.*, 2022b] Thiago S Gouvêa, Ilira Troshani, Marc Herrlich, and Daniel Sonntag. Interactive design of interpretable features for marine soundscape data annotation. In *Human-Centered Design of Symbiotic Hybrid Intelligence, HHAI*, Amsterdam, June 2022.

[Gunderson, 2008] L. Gunderson. Adaptive Management and Integrative Assessments. *Encyclopedia of Ecology*, pages 55–59, January 2008.

[Hartmann *et al.*, 2022] Mareike Hartmann, Han Du, Nils Feldhus, Ivana Kruijff-Korbayová, and Daniel Sonntag. XAINES: explaining AI with narratives. *Künstliche Intell.*, 36(3):287–296, 2022.

[Heinicke *et al.*, 2015] Stefanie Heinicke, Ammie K. Kalan, Oliver J.J. Wagner, Roger Mundry, Hanna Lukashevich, and Hjalmar S. Kühl. Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution*, 6(7):753–763, 2015.

[IPBES, 2019] IPBES. Global Assessment Report on Biodiversity and Ecosystem Services, May 2019.

[Jetz *et al.*, 2019] Walter Jetz, Melodie A. McGeoch, Robert Guralnick, Simon Ferrier, Jan Beck, Mark J. Costello, Miguel Fernandez, Gary N. Geller, Petr Keil, Cory Merow, Carsten Meyer, Frank E. Muller-Karger, Henrique M. Pereira, Eugenie C. Regan, Dirk S. Schmeller, and Eren Turak. Essential biodiversity variables for mapping and monitoring species populations. *Nat Ecol Evol*, 3(4):539–551, April 2019. 162 citations (Crossref) [2022-10-26] Number: 4 Publisher: Nature Publishing Group.

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

[Kath *et al.*, 2023a] Hannes Kath, Thiago S Gouvêa, and Daniel Sonntag. A deep generative model for interactive data annotation through direct manipulation in latent space. *arXiv:2305.15337*, 2023.

[Kath *et al.*, 2023b] Hannes Kath, Thiago S. Gouvêa, and Daniel Sonntag. A human-in-the-loop tool for annotating passive acoustic monitoring datasets. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence, IJCAI*, 2023.

[Kath *et al.*, 2023c] Hannes Kath, Bengt Lüers, Thiago S Gouvêa, and Daniel Sonntag. A virtual reality tool for representing, visualizing and updating deep learning models. *arXiv:2305.15353*, 2023.

[Kingma and Welling, 2013] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv:1312.6114*, 2013.

[Kingma *et al.*, 2014] Durk P Kingma, Shakir Mohamed, Danilo Jimenez Rezende, and Max Welling. Semi-supervised Learning with Deep Generative Models. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014.

[Parasuraman *et al.*, 2000] R. Parasuraman, T.B. Sheridan, and C.D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 30(3):286–297, May 2000. Conference Name: IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans.

[Pereira *et al.*, 2013] Henrique Miguel Pereira, Simon Ferrier, Michele Walters, Gary N Geller, Rob HG Jongman, Robert J Scholes, Michael William Bruford, Neil Brummitt, Stuart HM Butchart, AC Cardoso, et al. Essential biodiversity variables. *Science*, 339(6117):277–278, 2013.

[Plumptre, 2000] Andrew J Plumptre. Monitoring mammal populations with line transect techniques in african forests. *Journal of Applied Ecology*, 37(2):356–368, 2000.

[Prange and Sonntag, 2021] Alexander Prange and Daniel Sonntag. A demonstrator for interactive image clustering and fine-tuning neural networks in virtual reality. In *KI 2021: Advances in Artificial Intelligence: 44th German Conference on AI, Virtual Event, September 27–October 1, 2021, Proceedings 44*, pages 194–203. Springer, 2021.

[Ratner *et al.*, 2016] Alexander J Ratner, Christopher M De Sa, Sen Wu, Daniel Selsam, and Christopher Ré. Data Programming: Creating Large Training Sets, Quickly. In *Advances in Neural Information Processing Systems*, volume 29. NeurIPS, 2016.

[Rudin, 2019] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat Mach Intell*, 1(5):206–215, May 2019. Number: 5 Publisher: Nature Publishing Group.

[Siddharth *et al.*, 2017] N. Siddharth, Brooks Paige, Jan-Willem van de Meent, Alban Desmaison, Noah D. Goodman, Pushmeet Kohli, Frank Wood, and Philip H. S. Torr. Learning Disentangled Representations with Semi-Supervised Deep Generative Models, November 2017. arXiv:1706.00400 [cs, stat].

[Simard *et al.*, 2017] Patrice Y Simard, Saleema Amershi, David M Chickering, Alicia Edelman Pelton, Soroush Ghorashi, Christopher Meek, Gonzalo Ramos, Jina Suh, Johan Verwey, Mo Wang, et al. Machine teaching: A new paradigm for building machine learning systems. *arXiv:1707.06742*, 2017.

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

[Stowell, 2022] Dan Stowell. Computational bioacoustics with deep learning: a review and roadmap. *PeerJ*, 10:e13152, 2022.

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

[Tusfiqur *et al.*, 2022] Hasan Md Tusfiqur, Duy M. H. Nguyen, Mai T. N. Truong, Triet A. Nguyen, Binh T. Nguyen, Michael Barz, Hans-Juergen Profitlich, Ngoc T. T. Than, Ngan Le, Pengtao Xie, and Daniel Sonntag. Drg-net: Interactive joint learning of multi-lesion segmentation and classification for diabetic retinopathy grading, 2022.

[Ulloa *et al.*, 2021] Juan Sebastián Ulloa, Sylvain Haupert, Juan Felipe Latorre, Thierry Aubin, and Jérôme Sueur. scikit-maad: An open-source and modular toolbox for quantitative soundscape analysis in Python. *Methods in Ecology and Evolution*, 12(12):2334–2340, 2021.

[UNESCO, 2019] UNESCO. World Network of Biosphere Reserves, October 2019.

[Van Zoelen *et al.*, 2023] Emma Van Zoelen, Tina Mioch, Mani Tajaddini, Christian Fleiner, Stefani Tsaneva, Pietro Camin, Thiago S. Gouvêa, Kim Baraka, Maaike H. T. de Boer, and Mark A. Neerincx. Developing team design patterns for hybrid intelligence systems. In *Proceedings of the 2nd International Conference on Hybrid Human-Machine Intelligence, HHAI*, 2023.

[Watson *et al.*, 2014] James E. M. Watson, Nigel Dudley, Daniel B. Segan, and Marc Hockings. The performance and potential of protected areas. *Nature*, 515(7525):67–73, November 2014. Number: 7525 Publisher: Nature Publishing Group.

[Wilkinson *et al.*, 2016] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.

[Woodall *et al.*, 2021] Lucy C. Woodall, Sheena Talma, Oliver Steeds, Paris Stefanoudis, Marie-May Jeremie-Muzungaile, and Alain de Comarmond. Co-development, co-production and co-dissemination of scientific research: a case study to demonstrate mutual benefits. *Biol. Lett.*, 17(4):rsbl.2020.0699, 20200699, April 2021.

[Zacharias *et al.*, 2018] Jan Zacharias, Michael Barz, and Daniel Sonntag. A survey on deep learning toolkits and libraries for intelligent user interfaces. *arXiv:1803.04818*, 2018.