Wild Data Treasures: Towards Sustainable Practices in Deep Learning for Wildlife Monitoring

Ilira Troshani
Thiago S. Gouvêa
Daniel Sonntag
Interactive Machine Learning, German Research Centre for Artificial Intelligence (DFKI)
Oldenburg, Germany
Carl von Ossietzky University of Oldenburg
Oldenburg, Germany

ABSTRACT

While data collection and annotation is crucial for training supervised machine learning models and improving their accuracy, it can be resource-intensive. In this paper, we propose a weakly supervised method to extract fine-grained information from existing weakly-annotated data accumulated over time and alleviate the need for collection and annotation of fresh data. We also integrate it in an interactive tool that facilitates training and annotation. Communities comprising ecologists and other domain experts can use it to train machine learning models to detect animal species and monitor wildlife in protected areas. Our method not only improves the extraction of information from coarse labels but also simplifies the process of annotating new data for experts.. By lowering the time and expertise barrier to data annotation, we also aim to encourage individuals with varying levels of expertise to participate more in citizen science and contribute to preserving ecosystems.

CCS CONCEPTS

• **Human-centered computing** → *Interactive systems and tools.*

KEYWORDS

data annotation, machine learning, biodiversity conservation, sustainability

ACM Reference Format:

Ilira Troshani, Thiago S. Gouvêa, and Daniel Sonntag. 2024. Wild Data Treasures: Towards Sustainable Practices in Deep Learning for Wildlife Monitoring. In *Proceedings of ACM CHI 2024 Workshop on Human-Centered Green Technology for Community-wide Carbon Reduction (Sustaining Scalable Sustainability '24).* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Biodiversity conservation is crucial to the mitigation of CO_2 emissions. Plants such as trees capture CO_2 from the atmosphere to synthesize energy essential for their growth, constituting an indirect carbon capturing method [8]. Some plants and trees rely on animal species such as bats for pollination [5]. Maintaining healthy ecosystems helps keep carbon sinks and increase resilience to climate change. Creating and monitoring protected areas to support biodiversity and preserve the natural habitat of many species contributes to global climate regulation and CO_2 emission reduction. Data quality and quantity are key factors in the success of wildlife monitoring methods. However, data collection and annotation is a costly process, and its underlying digital infrastructure such as data centers and storage facilities is related to high power consumption resulting in high CO_2 emissions. The International Energy Agency estimates that data centers and data transmission are responsible for 1% of CO_2 emissions [1]. With the affordability of data acquisition decreasing, the process of capturing acoustic data for biodiversity monitoring has become more accessible. However, the storage demands for certain types of audio data, like ultrasound recordings, remain considerable. [7]. Gibb et al.[7] discuss some of the ways to minimize storage for acoustic data, but they also point out that discarding data is actually undesirable. Archived collections of acoustic data recorded during decades present a valuable data source that can be leveraged to develop new methods for species and behavior identification. Such collections include the FNJV Collection (Fonoteca Neotropical Jacques Vielliard at the University of Campinas, Brazil), and the Macaulay Library (Cornell Lab of Ornithology).

While most approaches focus on improving model training to reduce CO_2 emissions related to AI, we aim to tackle the issue starting from the data acquisition step, to make it more sustainable. In this context, we pose the following research questions:

(1) Question 1: How can we use existing collections more efficiently and support ecologists in their efforts for biodiversity conservation?
(2) Question 2: How can we support individual and collective efforts to annotate data for biodiversity conservation and climate change mitigation?

To answer question 1, we propose developing methods that use existing data collections more efficiently by extracting more information from them. This is a concrete step forward towards making AI and specifically deep learning more sustainable. To answer question 2, we focus on citizen science which is a valuable means of acquiring data and support biodiversity monitoring objectives including supporting management of protected areas in the decision-making process, increasing awareness to ecological changes, engaging public in ecological issues, developing ecological knowledge etc. [16, 18, 23]. However, data quality highly influences its use to achieve biodiversity objectives and often introduces inference bias [11]. A combination of Human/Computer Intelligence

Sustaining Scalable Sustainability '24, May 11, 2024, Honolulu, USA
2024. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

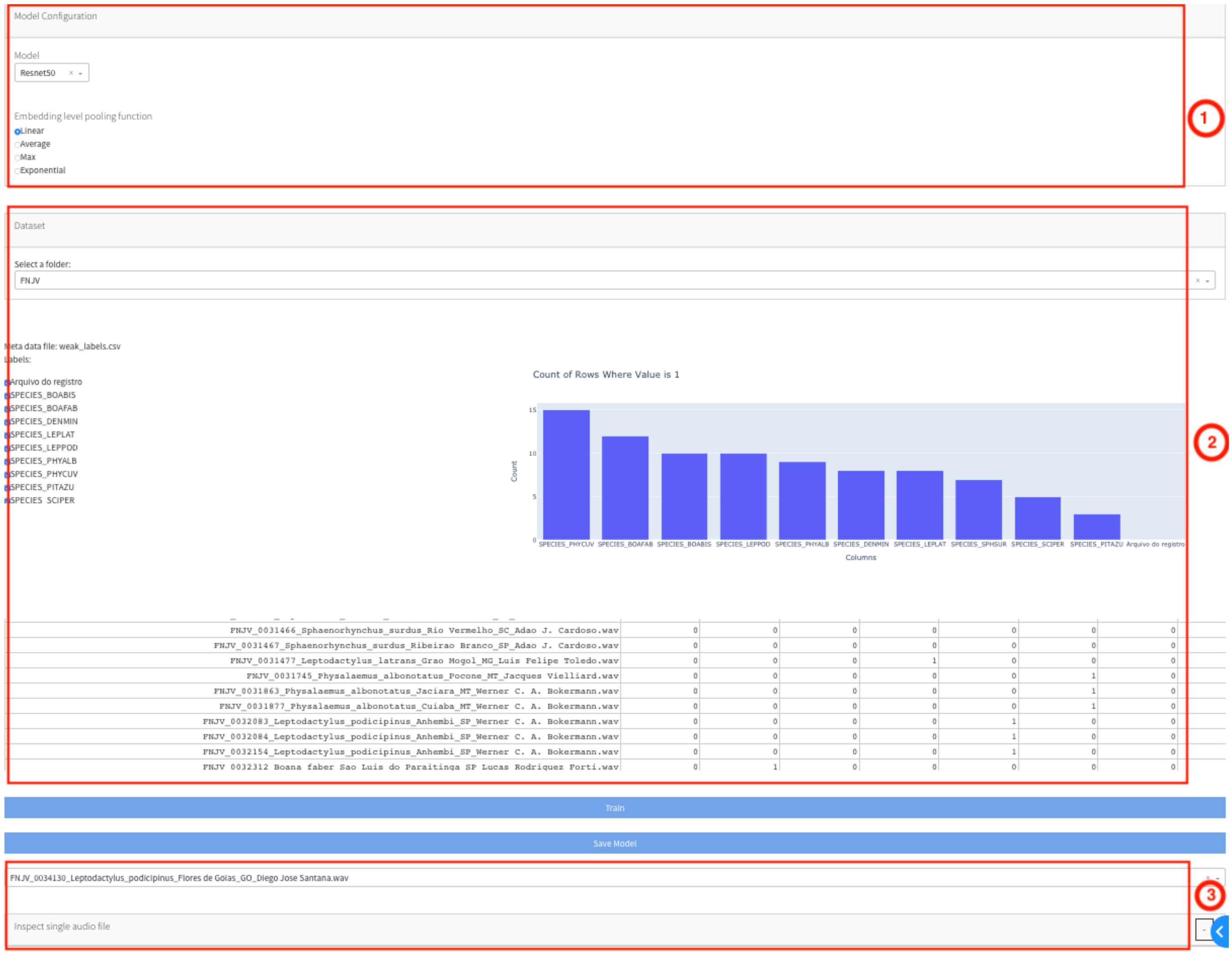


Figure 1: Training pane: Domain experts can train a model with a backbone of choice of pretrained models and configure parameters such as pooling function (model configuration area 1) using a weakly annotated dataset, such as a subset of the FNJV dataset for anurans. Upon selection of the dataset, the label distribution is computed for a better overview of the dataset (dataset area 2). At the bottom (single file inspection area 3), the user can inspect single spectrograms, here collapsed.

can be used to control data quality [21, 25]. By lowering the effort barrier to data annotation and automatically generating time intervals of the occurring species, we aim to increase participation in citizen science and improve the overall data quantity and quality.

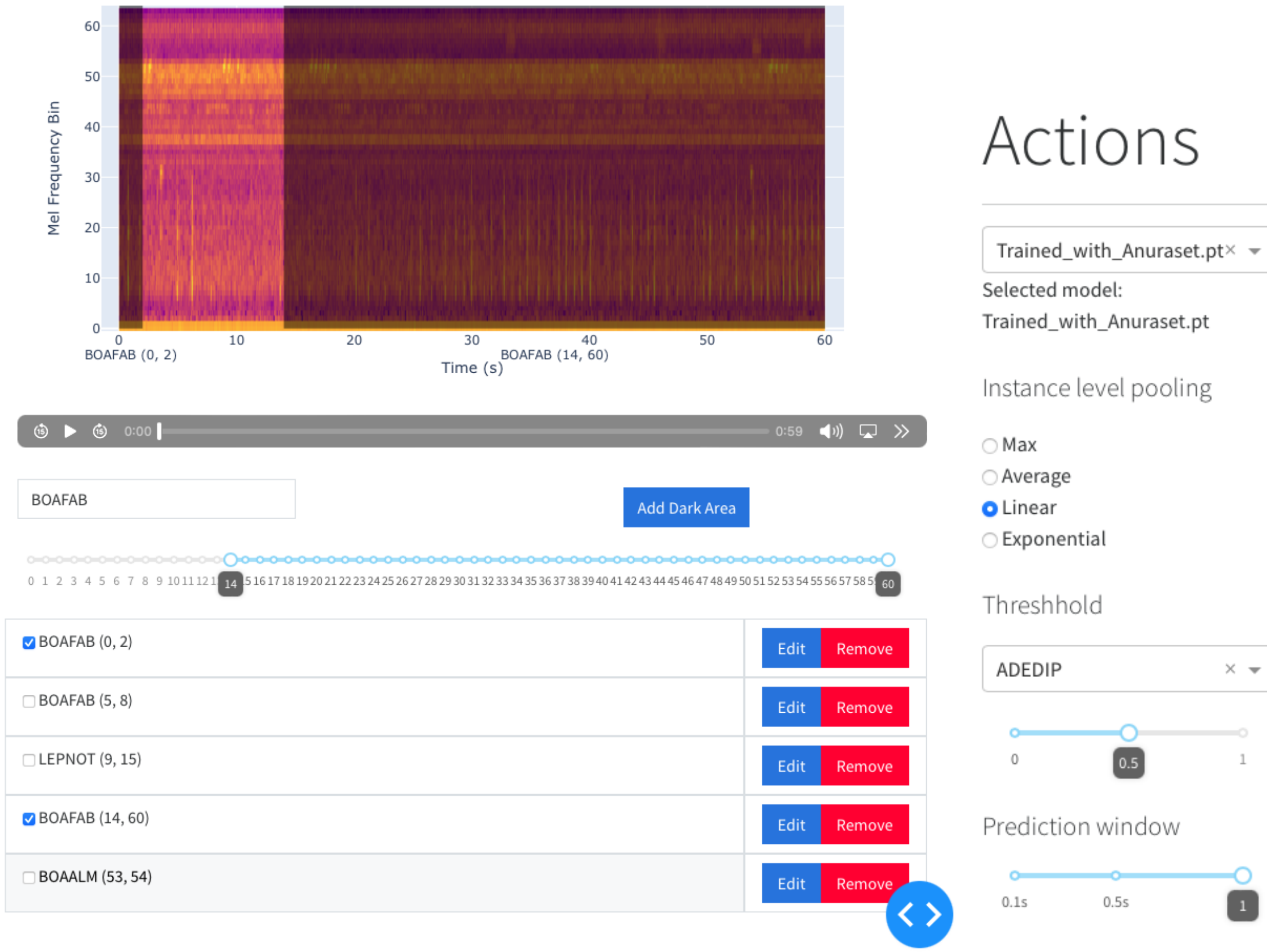
2 RELATED WORK

Deep learning methods have proven very useful for detection of sound events in animal sounds datasets. Some methods are associated with a user interface for result inspection and easy annotation [3, 4, 10]. A popular model for bird species classification is BirdNET [12]. Trained using weakly annotated datasets, in their method, the authors use heuristic image processing methods for signal-strength estimation [20]. Other approaches for training on weakly annotated data are weakly supervised learning and multiple instance learning [14]. Most methods developed for weakly supervised sound event detection (WSSED) use environment datasets such as AudioSet [6]. Some WSSED methods include [13, 15, 19, 24, 26, 27]. We contribute to this literature be exploring how WSSED can be used for wildlife monitoring to support biodiversity conservation and increase participation of individuals in citizen science.

3 METHOD

3.1 Model

In our approach, we use a Convolutional Recurrent Neural Network (CRNN) and Multiple Instance Learning (MIL) based on [26]. We train using AnuraSet [2], a benchmark dataset for animal sounds.



(a) Single File annotation

(b) WSL Settings

Figure 2: Annotation pane: Upon file upload, the model annotates the file with the detected species' calls and onset, offset time. The user can modify the predictions if necessary (a) using UI components such as slider, buttons and input fields. The more experienced machine learning practitioner can adjust the post-hoc parameters of the weakly supervised setting to achieve the best performance such as pooling function, prediction threshold and the precision of the prediction window for the onset-offset times (0.1 - 1s)(b).

To evaluate the method, we compute the F1 score for segment and file level predictions. We compare the performance of the CRNN trained from scratch with the feature extraction capabilities of models pretrained on visual data such as ResNet-50. We achieve similar performance with ResNet in both detection of events in file and segment level [22].

3.2 Interactive training and annotation tool

We design the tool with two user personas in mind: the expert and the novice, therefore we create a training (Figure 1) and an annotation pane (Figure 2a). The training pane has three main areas that are dynamically filled upon user action to minimize user input and collapsible for less visual clutter.

In the training pane, the expert can select a weakly-annotated dataset and a pretrained model, such as ResNet-50 [9]. As the pooling function plays an important role in the performance of weakly supervised learning models [22, 26] we make it a configurable parameter. In the annotation pane, the user can upload and annotate an audio file. Upon upload, inference runs in the background while the spectrogram of the audio file is computed and displayed. Upon selection, the time intervals are displayed as gray overlays on the spectrogram. The user can modify them, delete or add new ones. For the more experienced user we have created a Settings pane (Figure 2b) to modify post-hoc model parameters related to weakly supervised learning such as pooling function, prediction threshold to make the model more or less sensitive towards certain classes and prediction window. Currently, our tool is not meant for automatic dataset annotation, but to help novice and expert users, generate accurate annotations in a more efficient way. The application is implemented in Python using Dash [1] and PyTorch [17].

4 CONCLUSION AND FUTURE WORK

The purpose of this paper is to encourage sustainable ML methods and practices such as the use of existing data resources more efficiently, support biodiversity conservation efforts and help individuals get involved more in citizen science. Concretely, we proposed using museum data collections, contributing to decreasing the CO_2 footprint of machine learning. Considering the link between biodiversity conservation and CO_2 reduction, we proposed to scale biodiversity conservation efforts to individuals through citizen science by lowering the barrier to data annotation and increasing the accuracy of their contribution. We presented a method that extracts fine-grained information from weakly annotated museum collection data and a tool to train our model and create annotations for new data, addressing the needs of both expert and novice. Designed with the human "component" in mind, our tool makes model configuration and inference as transparent as possible and allows the user to modify the results if necessary. We plan to refine our tool by implementing more feedback-loops and include iterative model training.

REFERENCES

[1] International Energy Agency. 2024. Data Centres and Data Transmission Networks. https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

[2] Juan Sebastián Cañas, María Paula Toro-Gómez, Larissa Sayuri Moreira Sugai, Hernán Darío Benítez Restrepo, Jorge Rudas, Breyner Posso Bautista, Luís Felipe Toledo, Simone Dena, Adão Henrique Rosa Domingos, Franco Leandro de Souza, et al. 2023. A dataset for benchmarking Neotropical anuran calls identification in passive acoustic monitoring. *Scientific Data* 10, 1 (2023), 771.

[3] Kevin R Coffey, Ruby E Marx, and John F Neumaier. 2019. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. *Neuropsychopharmacology* 44, 5 (2019), 859–868.

[4] Yarden Cohen, David Aaron Nicholson, Alexa Sanchioni, Emily K Mallaber, Viktoriya Skidanova, and Timothy J Gardner. 2022. Automated annotation of birdsong with a neural network that segments spectrograms. *Elife* 11 (2022), e63853.

[5] Winifred F Frick, Tigga Kingston, and Jon Flanders. 2020. A review of the major threats and challenges to global bat conservation. *Annals of the New York Academy of Sciences* 1469, 1 (2020), 5–25.

[6] Jort F. Gemmeke, Daniel P. W. Ellis, Dylan Freedman, Aren Jansen, Wade Lawrence, R. Channing Moore, Manoj Plakal, and Marvin Ritter. 2017. Audio Set: An ontology and human-labeled dataset for audio events. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 776–780. https://doi.org/10.1109/ICASSP.2017.7952261

[7] Rory Gibb, Ella Browning, Paul Glover-Kapfer, and Kate E Jones. 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution* 10, 2 (2019), 169–185.

[8] Aysegul Yagmur Goren, Dogan Erdemir, and Ibrahim Dincer. 2024. Comprehensive review and assessment of carbon capturing methods and technologies: An environmental research. *Environmental Research* 240 (2024), 117503.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[10] Jia-jia Jiang, Ling-ran Bu, Fa-jie Duan, Xian-quan Wang, Wei Liu, Zhong-bo Sun, and Chun-yue Li. 2019. Whistle detection and classification for whales based on convolutional neural networks. *Applied Acoustics* 150 (2019), 169–178.

[11] Alison Johnston, Eleni Matechou, and Emily B Dennis. 2023. Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution* 14, 1 (2023), 103–116.

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

[13] Qiuqiang Kong, Yong Xu, and Mark D Plumbley. 2017. Joint detection and classification convolutional neural network on weakly labelled bird audio detection. In *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 1749–1753.

[14] Anurag Kumar and Bhiksha Raj. 2016. Audio Event Detection using Weakly Labeled Data. In *Proceedings of the 24th ACM international conference on Multimedia*. 1038–1047. https://doi.org/10.1145/2964284.2964310 arXiv:1605.02401 [cs].

[15] Koichi Miyazaki, Tatsuya Komatsu, Tomoki Hayashi, Shinji Watanabe, Tomoki Toda, and Kazuya Takeda. 2020. Weakly-Supervised Sound Event Detection with Self-Attention. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 66–70. https://doi.org/10.1109/ICASSP40776.2020.9053609 ISSN: 2379-190X.

[16] Gregory Newman, Mark Chandler, Malin Clyde, Bridie McGreavy, M Haklay, H Ballard, S Gray, R Scarpino, R Hauptfeld, D Mellor, et al. 2017. Leveraging the power of place in citizen science for effective conservation decision making. *Biological Conservation* 208 (2017), 55–64.

[17] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 8024–8035. http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[18] Hugh P Possingham, Brendan A Wintle, Richard A Fuller, and Liana N Joseph. 2012. The conservation return on investment from ecological monitoring. *Biodiversity monitoring in Australia* 49 (2012), 58.

[19] Ankit Shah, Anurag Kumar, Alexander G Hauptmann, and Bhiksha Raj. 2018. A closer look at weak label learning for audio events. *arXiv preprint arXiv:1804.09288* (2018).

[20] Elias Sprengel, Martin Jaggi, Yannic Kilcher, and Thomas Hofmann. 2016. Audio Based Bird Species Identification using Deep Learning Techniques. *LifeCLEF 2016* (2016).

[21] Brian L Sullivan, Jocelyn L Aycrigg, Jessie H Barry, Rick E Bonney, Nicholas Bruns, Caren B Cooper, Theo Damoulas, André A Dhondt, Tom Dietterich, Andrew Farnsworth, et al. 2014. The eBird enterprise: An integrated approach to development and application of citizen science. *Biological conservation* 169 (2014), 31–40.

[22] Ilira Troshani, Thiago Gouvea, and Daniel Sonntag. 2024. Leveraging Sound Collections for Animal Species Classification with Weakly Supervised Learning. In *3rd Annual AAAI Workshop on AI to Accelerate Science and Engineering. AAAI Workshop on AI to Accelerate Science and Engineering (AI2ASE-2024), AAAI, Vancouver, Canada*.

[23] Ayesha IT Tulloch, Hugh P Possingham, Liana N Joseph, Judit Szabo, and Tara G Martin. 2013. Realising the full potential of citizen science monitoring programs. *Biological Conservation* 165 (2013), 128–138.

[24] Panagiotis Tzirakis, Alexander Shiarella, Robert Ewers, and Björn W Schuller. 2020. Computer audition for continuous rainforest occupancy monitoring: the case of Bornean gibbons' call detection. (2020).

[25] 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. 2023. Developing Team Design Patterns for Hybrid Intelligence Systems. In *Frontiers in Artificial Intelligence and Applications*, Paul Lukowicz, Sven Mayer, Janin Koch, John Shawe-Taylor, and Ilaria Tiddi (Eds.). IOS Press. https://doi.org/10.3233/FAIA230071

[26] Yun Wang, Juncheng Li, and Florian Metze. 2019. A Comparison of Five Multiple Instance Learning Pooling Functions for Sound Event Detection with Weak Labeling. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 31–35. https://doi.org/10.1109/ICASSP.2019.8682847 ISSN: 2379-190X.

[27] Jie Xie, Kai Hu, Mingying Zhu, and Ya Guo. 2020. Bioacoustic signal classification in continuous recordings: Syllable-segmentation vs sliding-window. *Expert Systems with Applications* 152 (2020), 113390.

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