forum acusticum 2023

AUTOMATIC DETECTION OF INDRIS' SONGS USING CONVOLUTIONAL NEURAL NETWORKS

Valente Daria[1*] **Ravaglia Davide**[1] **Ferrario Valeria**[2] **De Gregorio Chiara**[1] **Carugati Filippo**[1] **Raimondi Teresa**[1] **Cristiano Walter**[1] **Torti Valeria**[1] **Von Hardenberg Achaz**[2] **Ratsimbazafy Jonah**[3] **Giacoma Cristina**[1] **Gamba Marco**[1]

[1] Department of Life Sciences and Systems Biology, University of, Torino, Italy
[2] Conservation Biology Research Group, University of Chester, UK
[3] GERP, Fort Duchesne, Antananarivo 101, Madagascar, 101 Antananarivo, Madagascar

ABSTRACT

The combination of bioacoustic monitoring with machine learning algorithms is increasingly used to gain information on species distribution or activity. Among these methods, convolutional neural networks (CNN) for the automatic classification of both environmental and animal sounds proved particularly effective. We employed an automated classifier based on a CNN aimed at detecting the presence of *Indri indri* songs recorded in Maromizaha Forest from 2019 to 2022 via passive acoustic monitoring. The network achieved high accuracy (>90%) and recall (>80%) values in assessing the songs presence while the use of data augmentation and transfer learning was able to generalize to unsampled periods. Lastly, our process was able to correctly describe both daily and annual pattern of indris' singing behavior, critical piece of information to plan data collection and conservation practices.

Keywords: CNN - passive acoustic monitoring - *Indri indri* - singing primates

[*] ***Corresponding author***: *daria.valente@unito.it*

1. INTRODUCTION

Passive acoustics is nowadays a well-established tool for biodiversity monitoring, and for gaining broad-scale and high-resolution acoustic time series [1]. Long-term studies relying on passive acoustic monitoring (PAM) allow investigating temporal and spatial dynamics at an unprecedented level of detail but are also generating enormous amount of data. The use of machine learning to ease the processing of data collected through passive acoustic monitoring is becoming a common procedure in bioacoustic research, at least for what concerns the reduction of the amount of data to be manually processed or labelled (e.g. [2]). Still, the potential for applying these algorithms to answer ecologically significant questions remains unexploited. Our aim was to use machine learning algorithms to extract ecologically relevant information. We first aimed at understanding whether a convolutional neural network could be used to detect the vocalisations of a Critically Endangered singing primate (*Indri indri*) within recordings collected via passive acoustics, at both temporal and spatial scale. Indeed, the vocal pattern of many taxa peaks at dawn or dusk and shows diel variations; considering a longer span, vocal behavior may show a relevant seasonality. Investigating these variations may be useful to gain insights into the spatial and temporal ecology of a species, as well as to understand potential fluctuations of its vocal behavior in relation to external factors.

2. CNN APPLIED TO INDRIS' SONGS

We collected data in Maromizaha, a pristine rainforest patch located in northeastern Madagascar, a region with average annual precipitation comprised between 1500 mm and 2000 mm, and temperatures spanning between 15°C and 23°C. 66,542 10-minute recordings have been recorded in a three-year span (2019 through 2021) using 12 ARUs set (autonomous recording units; Figure 1) set to record 10 minutes every 30 minutes across 24 hours (48 recordings per day), all-year long. Recordings were distributed across years as follows. AudioMoth: 28,314 files recorded in 2019 and 12,661 recorded in 2020/21 (respectively AM2019 and AM 2020/21); Song Meter SM4: 25,468 files recorded in 2020/21 (SM4 2020/2021).

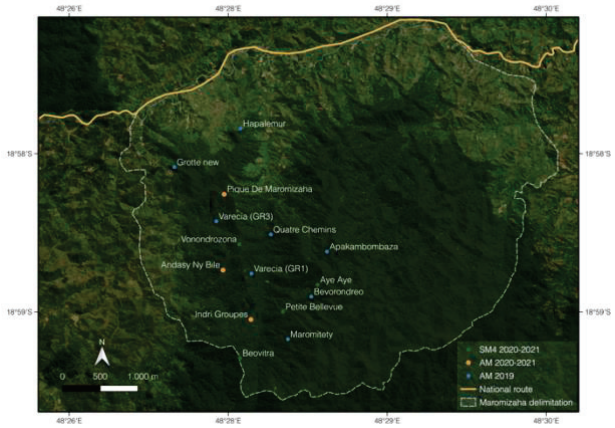

Figure 1. The map shows the territorial boundaries of as well as the location of the sites where we positioned the ARUs within the forest. We used two types of ARUs: ten AudioMoth (AM, in blue and yellow) and two Wildlife Acoustics Song Meter SM4 recorders (SM, in green). The sampled area covered about 700 hectares.

We first aimed at verifying whether indris' presence can be automatically detected from passively recorded data trough machine learning algorithms. We therefore focused on a species-specific loud call, the song, able to span several kilometers from the source of emission [3] and easily distinguishable within passively recorded data [4]. Despite indris' songs covering a wide frequency range, with the upper harmonics above 15 kHz, for the manual selection, we decided to focus on a narrow range, up to 2500 Hz. This portion indeed includes the crucial portion of the songs, i.e. the fundamental frequency and the first harmonics. We first converted the .wav files into spectrograms, annotating absence and presence of songs (files containing songs accounted for the 10% of the sample, on average). To choose which features to use to feed the network (i.e. images or acoustic parameters, on a linear or logarithmic scale), we first verified that considering the maximum frequency range of the spectrum (0-20 kHz) and a logarithmic scale sensibly improved the performance (Figure 2).

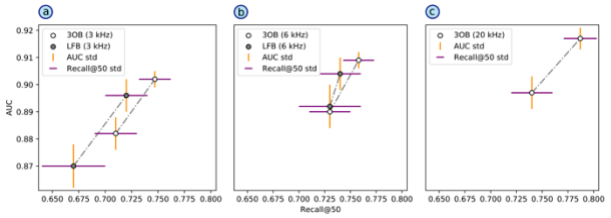

Figure 2. Values of AUC (Recall and Fall-out) and Recall@50 (Recall when Precision equals 0.5) shown for linear (LFB, linear frequency bins) and logarithmic (3OB, third-octave bands) frequency bands covering 3 kHz (a), 6 kHz (b), and 20 kHz (c). In the latter case we only computed the third-octave bands to keep the number of features contained. For each couple of points connected by the dashed line, the one with better performance was obtained by applying data augmentation.

We therefore transformed each file into a matrix of features 30 time intervals × 18 frequency bands (up to 20 kHz) based on a third-octave bands system. We then built a six-layer Convolutional Neural Network [5]. The output of these layers is flattened into 240 neurons. From now on the networks becomes a fully-connected multilayer perceptron. We then applied a dropout between the flattened layer and an additional one of 200 neurons. Here we applied five more neurons indicating information regarding time of day and week of recording (two neurons codifying the hour, two codifying the week, one resulting from the four combined together). These 205 neurons directly lead to a binary output (presence or absence of song).

We fine-tuned the CNN architecture by using the 70% of the AM2019 dataset as training set; the remaining 20%

and 10% were used as validation and test set, respectively. We then saved the model and the relative weights obtained from the training on the whole set AM2019. Starting from this model, we fine-tuned the transfer learning procedure by using both 2020/21 sets (AM and SM) dividing them into a training (5%), validation (15%), and test set (80%). Once proved the efficacy of our network (see Table 1), we included two new sets of data: 13,490 files recorded in 2021 and 2022 via two Song Meters SM4 (SM4 2021-2022); 10,680 files recorded in 2022 via three Song Meter Micro (SMM 2022). We first manually labeled the 10% of the recordings and then equally divided it into a training and test set. We did not use a validation test because we did not modify the network structure. Lastly, we re-trained the model by using the entire set of data (AM 2019, AM 2020-2021, SM4 2020-2021) and found comparable results when using the 5% of each of the two new sets as test sets. We obtained the following metrics. SM4 2021-2022: Accuracy = 0.93, Precision = 0.61, Recall = 0.84; SMM 2022: Accuracy = 0.90, Precision = 0.55, Recall = 0.77. These values corroborate the robustness of our architecture, which proved to be consistent even on new data, without the need of being constantly fine-tuned.

From an ecological perspective, we found our network to be able i) to reliably detect the presence of indris' songs; ii) to do so consistently across years and recording sites; iii) to correctly describe seasonal and circadian pattern of indris' singing behavior. In fact, we found the lowest probability of detecting songs during the cool dry season (May - September) and the highest probability during the warm rainy season (October - April), with a steady rise between the end of October and the end of December. When investigating the daily singing pattern, we found an increase in the calling activity rapidly following the astronomical sunrise, and a peak comprised between 6 and 10 am.

	Test set							
	AM 2019		AM 2020/2021			SM4 2020/2021		
TL	-		-	+		-	+	
DA	-	+	-		+	-		+
Accuracy	0.94 ± 0.01	0.90 ± 0.01	0.89 ± 0.01	0.96 ± 0.01	0.92 ± 0.02	0.92 ± 0.01	0.95 ± 0.01	0.91 ± 0.01
Precision	0.85 ± 0.04	0.59 ± 0.03	0.44 ± 0.01	0.81 ± 0.04	0.56 ± 0.04	0.59 ± 0.01	0.83 ± 0.02	0.52 ± 0.04
Recall	0.69 ± 0.02	0.82 ± 0.04	0.89 ± 0.02	0.71 ± 0.03	0.86 ± 0.02	0.75 ± 0.01	0.63 ± 0.03	0.84 ± 0.02
F1 Score	0.76 ± 0.02	0.69 ± 0.01	0.59 ± 0.01	0.76 ± 0.01	0.68 ± 0.02	0.66 ± 0.01	0.71 ± 0.01	0.64 ± 0.03

Table 2. Accuracy, Precision, Recall, and F1 score obtained for the test sets of the three sets of data (AM 2019; AM 2020/2021; and SM4 2020/2021); the uncertainty associated with the different measures was obtained through a 10-fold cross-validation. This procedure was performed alongside the transfer learning for the two 2020/2021 test sets. We report the variation in the metrics before and after data augmentation (DA) and/or transfer learning (TL).

3. CONCLUSION

Our method was able to ease the automated detection of vocalizations within a large set of data recorded trough autonomous recording units, by reducing the amount of data to analyse and therefore by minimizing both computational time and storage space. Our results show a slightly worse performance compared to similar works (i.e. studies on sperm whale echolocation clicks [6] and bat [7] vocalisations accuracy values of 99.5% and 99.7%, respectively. However, these studies relied on the classification of spectrograms [6] or images reconstructed from Mel-frequency cepstral coefficients [7]. In contrast, we built a network relying on matrices of features up to 10,000 times lighter than the original waveform. On the one hand, this ensures the acceleration of whole entire detection procedure and enables more frequent data exchange between research groups. On the other hand the data reduction can, in turn, cause a decrease in the preserved information and by consequence impact the algorithm's recognition

capacity. However, our process was able to depict ecologically relevant information, such as the pattern of indris' singing behavior both at a short and long-temporal scale, and possibly to provide a glimpse into the spatial ecology of the species. Indeed, we found the song occurrence to be inversely related to the conservation status of the forest: the more degraded the site, the lower the amount of recorded of songs. This is in line with a recent study showing indris to be edge-intolerant [8]. Concluding, our process was able to contribute significantly to the automated detection of loud-calling species as well as to provide essential information to plan data collection and conservation policies, critical to study and protect endangered species and the fragile environment where they live.

4. ACKNOWLEDGMENTS

We wish to express our gratitude to the local research guides that helped during the data collection. We are also grateful to the GERP (Groupe d'Étude et de Recherche sur les Primates de Madagascar) and the Parco Natura Viva for the precious support during the research activities. We thank the Ministère de l'Environnement et du Développement Durable (MEDD) for granting the permits for this study (118/19/MEDD/SG/DGEF/DSAP/DGRNE, 284/19/MEDD/SG/DGEF/DSAP/DGRNE, 338/19/MEDD/G/DGEF/DSAP/DGRNE, 186/22/MEDD/SG/DGGE/DAPRNE/SCBE.Re).

5. REFERENCES

[1] SJ. Ross, DP. O'Connell, JL. Deichmann, C. Desjonquères, A. Gasc, JN. Phillips, SS. Sethi, CM. Wood, Z. Burivalova, "Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions." *Functional Ecology*, 37, pp. 959–975, 2023.

[2] E. Dufourq, I. Durbach, JP. Hansford, A. Hoepfner, H. Ma, JV, Bryant, CS. Stender, W. Li, Z. Liu, Q. Chen Q. Automated Detection of Hainan Gibbon Calls for Passive Acoustic Monitoring. *Remote Sensing in Ecology and Conservation*, 7, pp.475–487, 2021.

[3] JI. Pollock. The Song of the Indris (*Indri indri*; Primates: Lemuroidea): Natural History, Form, and Function. *International Journal of Primatology* 7, pp. 225-264, 1986.

[4] E. Browning, R. Gibb, P. Glover-Kapfer, KE. Jones. *Passive Acoustic Monitoring in Ecology and Conservation*; WWF-UK: Woking, UK, 2017.

[5] JT. Springenberg, A. Dosovitskiy T. Brox T, M. Riedmiller. Striving for simplicity: The all convolutional net. arXiv:1412.6806, 2014.

[6] PC. Bermant, MM. Bronstein, RJ. Wood, S. Gero, DF. Gruber, Deep Machine Learning Techniques for the Detection and Classification of Sperm Whale Bioacoustics. *Scientific Reports* 9, 12588, 2019.

[7] Y. Paumen, M. Mälzer, S. Alipek, J. Moll, B. Lüdtke, H. Schauer-Weisshahn. Development and Test of a Bat Calls Detection and Classification Method Based on Convolutional Neural Networks. *Bioacoustics* 31, pp. 505–516, 2022.

[8] T. King, R. Dolch, HNT. Randriahaingo, L. Randrianarimanana, M. Ravaloharimanitra, *Indri indri*. The IUCN Red List of Threatened Species. 2020.