





# An integrated passive acoustic monitoring and deep learning pipeline for black-and-white ruffed lemurs (*Varecia variegata*) in Ranomafana National Park, Madagascar

Carly H. Batist<sup>1,2,3</sup>  | Emmanuel Dufourq<sup>4,5,6,7</sup>  | Lorène Jeantet<sup>4,5,6</sup>  |  
Mendrika N. Razafindraibe<sup>8,9</sup> | Francois Randriamanantena<sup>10</sup> | Andrea L. Baden<sup>1,2,11</sup> 

<sup>1</sup>Department of Anthropology, City University of New York (CUNY) Graduate Center, New York, New York, USA

<sup>2</sup>New York Consortium in Evolutionary Primatology (NYCEP), New York, New York, USA

<sup>3</sup>Rainforest Connection (RFCx), Katy, Texas, USA

<sup>4</sup>African Institute for Mathematical Sciences, Muizenberg, South Africa

<sup>5</sup>Department of Mathematical Sciences, Stellenbosch University, Stellenbosch, South Africa

<sup>6</sup>National Institute for Theoretical & Computational Sciences, Stellenbosch, South Africa

<sup>7</sup>African Institute for Mathematical Sciences, Research and Innovation Centre, Kigali, Rwanda

<sup>8</sup>Department of Animal Biology, University of Antananarivo, Antananarivo, Madagascar

<sup>9</sup>Institut International de Science Sociale, Antananarivo, Madagascar

<sup>10</sup>Centre ValBio, Ranomafana, Madagascar

<sup>11</sup>Department of Anthropology, Hunter College of City University of New York (CUNY), New York, New York, USA

## Correspondence

Carly H. Batist, CUNY Graduate Center, 365  
5th Ave, New York, NY 10016, USA.  
Email: [cbatist@gradcenter.cuny.edu](mailto:cbatist@gradcenter.cuny.edu)

## Funding information

Graduate Center; International Development  
Research Centre; Global Affairs Canada;  
African Institute for Mathematical Sciences,  
South Africa

## Abstract

The urgent need for effective wildlife monitoring solutions in the face of global biodiversity loss has resulted in the emergence of conservation technologies such as passive acoustic monitoring (PAM). While PAM has been extensively used for marine mammals, birds, and bats, its application to primates is limited. Black-and-white ruffed lemurs (*Varecia variegata*) are a promising species to test PAM with due to their distinctive and loud roar-shrieks. Furthermore, these lemurs are challenging to monitor via traditional methods due to their fragmented and often unpredictable distribution in Madagascar's dense eastern rainforests. Our goal in this study was to develop a machine learning pipeline for automated call detection from PAM data, compare the effectiveness of PAM versus in-person observations, and investigate diel patterns in lemur vocal behavior. We did this study at Mangevo, Ranomafana National Park by concurrently conducting focal follows and deploying autonomous recorders in May–July 2019. We used transfer learning to build a convolutional neural network (optimized for recall) that automated the detection of lemur calls (57-h runtime; recall = 0.94, F1 = 0.70). We found that PAM outperformed in-person

**Abbreviations:** CNN, convolutional neural network; ML, machine learning; PAM, passive acoustic monitoring; SM4, SongMeter4.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *American Journal of Primatology* published by Wiley Periodicals LLC.

observations, saving time, money, and labor while also providing re-analyzable data. Using PAM yielded novel insights into *V. variegata* diel vocal patterns; we present the first published evidence of nocturnal calling. We developed a graphic user interface and open-sourced data and code, to serve as a resource for primatologists interested in implementing PAM and machine learning. By leveraging the potential of this pipeline, we can address the urgent need for effective primate population surveys to inform conservation strategies.

#### KEYWORDS

conservation, convolutional neural network, ecoacoustics, machine learning, strepsirrhine

## 1 | INTRODUCTION

Given the current biodiversity loss crisis, there is a growing need for cost-effective and scalable wildlife monitoring solutions. Passive acoustic monitoring (PAM) is an increasingly popular tool in which specialized sound recorders are deployed across a study area to record sound autonomously using specified configuration parameters (Blumstein et al., 2011; Gibb et al., 2019; Piel et al., 2021). These recorders can capture species-specific calls, which can later be modeled to estimate species presence, occupancy, and distribution (Gibb et al., 2019; Sugai et al., 2019). PAM can detect some species up to 10x more frequently than camera-traps due to the omnidirectional nature of microphones and resultant wider detection area (Crunchant et al., 2020; Enari et al., 2019), and can be as accurate and efficient as human observation (Boullhesen et al., 2021; Digby et al., 2013; Leach et al., 2016), if not more so (Castro et al., 2019; Darras et al., 2019; Hoefer et al., 2023; Melo et al., 2021). PAM is especially powerful because it is noninvasive, cost-efficient, and able to yield long-term data that can be objectively and repeatedly analyzed (Blumstein et al., 2011; Gibb et al., 2019; Kvsn et al., 2020).

PAM and automated detection models have a long history of use with marine mammals, making this taxonomic subfield a pioneer in PAM research (Bittle & Duncan, 2013; Mellinger et al., 2007; Zimmer, 2011). However, the use of PAM with nonflying terrestrial mammals has been comparatively less frequent, accounting for only 6% of studies in a recent meta-analysis (28 of 460; Sugai et al., 2019). More recently, however, PAM has become widely employed for birds (Zhong et al., 2021), bats (Paumen et al., 2022), and anurans (Melo et al., 2021). Nevertheless, PAM remains underrepresented in the primatology literature (Table 1). This is particularly surprising, given that the Order contains several highly vocal taxa (Table 1). Among these are Madagascar's lemurs, considered the world's most endangered mammals (Schwitzer et al., 2014). Remarkably, despite Madagascar's status as a biodiversity hotspot and the country's distressing loss of half its forests within the last six decades (Vieilledent et al., 2018), there has been a notable lack of PAM studies in this region. Given the potential contributions of PAM toward conservation monitoring efforts in Madagascar, our study seeks to address this gap by investigating the potential of PAM to survey lemur populations and inform conservation initiatives.

The conservation of the Critically Endangered black-and-white ruffed lemurs (*Varecia variegata*; hereafter ruffed lemurs) is especially important given their vital roles as seed dispersers and pollinators; some fruiting trees rely entirely on ruffed lemurs as their sole dispersers (Britt, 2000; Dew & Wright, 1998; Federman et al., 2016; Moses & Semple, 2011). Additionally, ruffed lemurs are very sensitive to disturbed habitat and can forego breeding up to 4 or more years when environmental conditions are not suitable (Baden et al., 2013; Ratsimbazafy, 2002; Vasey & Borgerson, 2009). Ruffed lemurs are therefore both indicators and determinants of rainforest health.

Ruffed lemurs are limited to fragments of the once-continuous eastern Malagasy rainforests (Louis et al., 2020). They are high-canopy specialists (Beeby & Baden, 2021) and can be difficult for observers on the ground to find and continuously track. Ruffed lemurs' patchy distribution throughout a mosaic of forest fragments makes it difficult to accurately assess population distribution or size (Irwin et al., 2005; Louis et al., 2020; Morelli et al., 2020). PAM is particularly useful in such dense, montane forest habitats because it enables animal detection in areas with limited visibility and arduous topography (Gibb et al., 2019; Kalan et al., 2015; Thompson et al., 2010).

Ruffed lemurs emit distinct choruses (roar-shrieks and barks) that make them an ideal candidate species for PAM (Batist et al., 2022; Pereira et al., 1988; Petter & Charles-Dominique, 1979; Turner & Harrenstien, 1985). Roar-shrieks and barks are considered ruffed lemur long calls and travel the furthest, making them effective with acoustic monitoring (Batist et al., 2022, 2023; Pereira et al., 1988; Petter & Charles-Dominique, 1979; Turner & Harrenstien, 1985). Acoustic features of these calls are listed in Table 2 and spectrograms are provided in Figure 1 (though from a directional microphone during active acoustic monitoring). These calls are given frequently, but not at any particular time of day (Batist et al., 2022; Geissmann & Mutschler, 2006). They are multisyllabic contagious choruses, meaning that once one individual within a subgroup begins calling, the rest of the subgroup joins in (Pereira et al., 1988; Turner & Harrenstien, 1985). Roar-shrieks in one subgroup are often answered by other neighboring, out-of-sight subgroups as well. The exact function of roar-shrieks is debated, but it seems to be a sort of roll-call among individuals following potentially stressful situations (Batist et al., 2022; Pereira et al., 1988;

**TABLE 1** Passive acoustic monitoring studies with primates (compiled from Google Scholar and Web of Science searches in May 2023 with combinations of “PAM” or “passive acoustic monitoring” AND “primate” or “monkey” or “ape” or “lemur”).

Paper	Site(s)	Species/taxon	Use case/research aim	DL?
Bergler et al. (2022)	Not listed	Chimpanzee, mandrill, red-capped mangabey, guenon (mixed-species)	Developing/comparing deep learning classification models	Yes
Bittencourt et al. (2023)	Estação Ambiental de Peti, Brazil	Black-tufted marmosets ( <i>Callithrix penicillata</i> )	Impact of mining activity on marmoset vocal activity	No
Clink nad Klink (2021)	Malaysian Borneo	Northern gray gibbon ( <i>Hylobates funereus</i> )	Comparing unsupervised clustering algorithms, individual identification	No (CL)
Crunchant et al. (2020)	Issa Valley, Tanzania	Chimpanzee ( <i>Pan troglodytes</i> )	Comparing detection and occupancy vs. camera-trapping	No
Do Nascimento et al. (2024)	Virua National Park, Brazil	Guianan red howler monkey ( <i>Alouatta macconnelli</i> )	Diel patterns and acoustic structure	No (TM)
Dufourq et al. (2021)	Hainan, China	Gibbon ( <i>Nomascus hainanus</i> )	Developing deep learning detection model	Yes
Enari et al. (2019)	Shirakami, Asahi, Japan	Japanese macaque ( <i>Macaca fuscata</i> )	Comparing detection with camera traps	No (HMM)
Heinicke et al. (2015)	Tai, Côte d'Ivoire	Chimpanzee ( <i>Pan troglodytes</i> ), Diana monkey ( <i>Cercopithecus diana</i> ), king colobus ( <i>Procolobus badius</i> ), red colobus	Developing automated detection model	No (SYM + GMM)
Hutschenreiter et al. (2022)	Riviera Maya, Yucatán peninsula, Mexico	Geoffroy's spider monkey ( <i>Ateles geoffroyi</i> )	Assess influence of land-use and disturbance	No
Kalan et al. (2015)	Tai, Côte d'Ivoire	Chimpanzee ( <i>Pan troglodytes</i> ), Diana monkey ( <i>Cercopithecus diana</i> ), king colobus ( <i>Procolobus badius</i> ), red colobus	Occupancy estimation/models	No (SYM + GMM)
Kalan et al. (2016)	Issa Valley, Tanzania & Tai, Côte d'Ivoire	Chimpanzee ( <i>Pan troglodytes</i> )	Ranging behavior and habitat use	No (SYM + GMM)
Lawson et al. (2023)	Osa peninsula, Costa Rica	Geoffroy's spider monkey ( <i>Ateles geoffroyi</i> )	Occupancy models/GLMs to assess influence of land-use and disturbance	Yes
Markolf et al. (2022)	Kirindy Forest, Madagascar	Pale fork-marked lemur ( <i>Phaner pallescens</i> )	Estimate vocal activity and density	No
Pérez-Granados and Schuchmann (2021)	SESC Pantanal	Black and gold howler monkey ( <i>Alouatta caraya</i> )	Assessing temporal patterns (diel, seasonal, annual)	No
Piel (2018)	Issa Valley, Tanzania	Chimpanzee ( <i>Pan troglodytes</i> )	Diel activity, temporal patterns	No
Ravaglia et al. (2023)	Maromizaha, Madagascar	Indri ( <i>Indri indri</i> )	Automated detection model	Yes
Spillmann et al. (2015)	Tuanan, Central Kalimantan, Indonesia	Bornean orangutan ( <i>Pongo pygmaeus</i> )	Acoustic localization	No (HMM)
Tzirakis et al. (2020)	Malaysian Borneo	Bornean gibbon ( <i>Hylobates muelleri</i> )	Developing CNN for call detection	Yes
Vu and Tran (2019)	Cat Tien National Park, Vietnam	Southern yellow-cheeked crested gibbon ( <i>Nomascus gabriellae</i> )	Estimating probability of occurrence in different areas	No

(Continues)

TABLE 1 (Continued)

Paper	Site(s)	Species/taxon	Use case/research aim	DL?
Wang, Wen, et al. (2022)	Hainan, China	Gibbon ( <i>Nomascus hainanus</i> )	Developing deep learning detection model	Yes
Wang, Ye, et al. (2022)	Hainan, China	Gibbon ( <i>Nomascus hainanus</i> )	Comparing deep learning models with variable calls	Yes
Wood et al. (2023)	Southeastern Chiapas, Mexico	Yucatán black howler monkey ( <i>Alouatta pigra</i> )	Extending BirdNET to non-bird taxa, using occupancy models	Yes
Zaluar et al. (2022)	Tijuca Forest, Rio de Janeiro, Brazil	Marmosets ( <i>Callithrix jacchus</i> , <i>C. penicillata</i> )	Impact of invasive marmosets on bird diversity	No
Zambolli et al. (2023)	Santa Maria, Pontal do Paranapanema, Brazil	Black lion tamarin ( <i>Leontopithecus chrysopygus</i> )	Determine sampling effort to confirm species presence	No (TM)
Zwerts et al. (2021)	Mefou, Cameroon (semi-captive)	Chimpanzee, mandrill, red-capped mangabey, guenon (mixed-species)	Building training data set for machine learning models	Yes

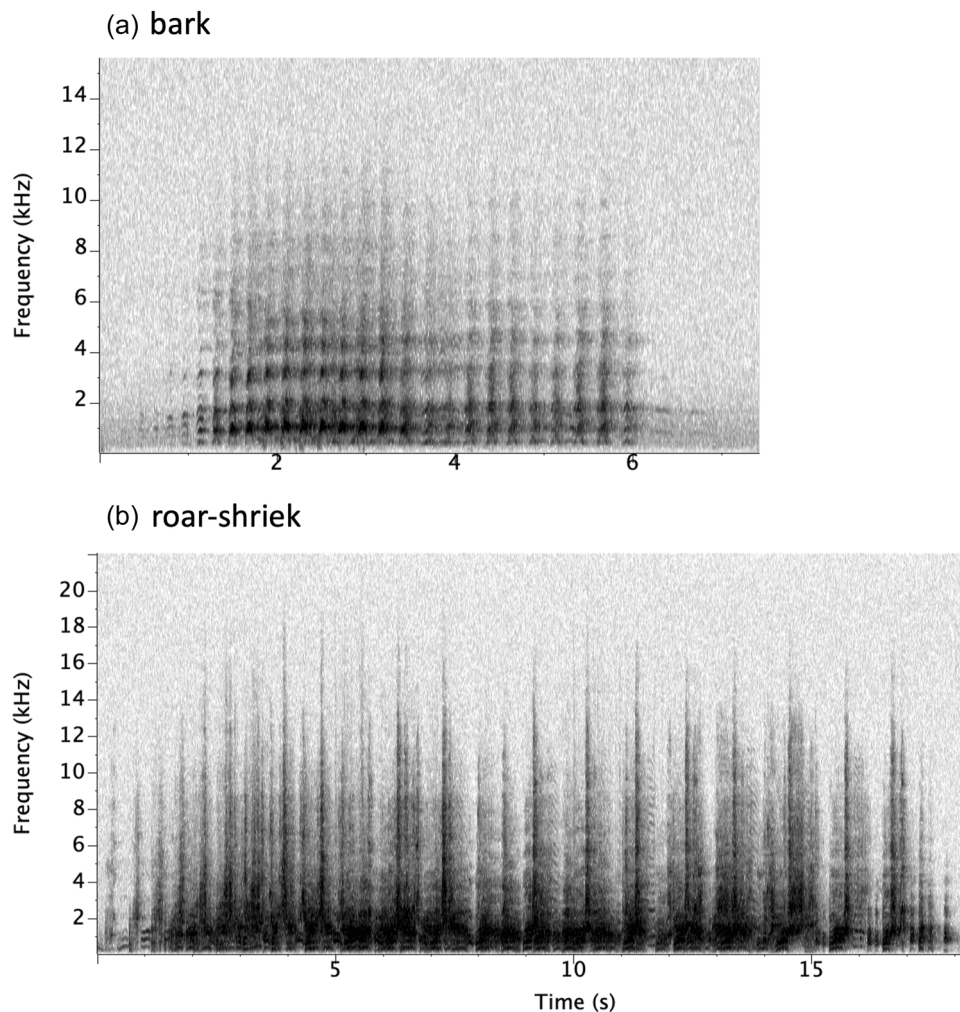
Abbreviations: CL, clustering; DL, deep learning; GMM, Gaussian mixture model; HMM, hidden Markov model; SVM, support vector machine; TM, template matching.

TABLE 2 Acoustic features of the black-and-white ruffed lemur roar-shriek and bark in Mangevo, Ranomafana National Park, Madagascar (adapted from Batist et al., 2023).

Acoustic feature (units)	Roar-shriek	Bark
Mean frequency (kHz)	2.84 ± 0.47	1.78 ± 0.50
Inter-quartile range (frequency; kHz)	2.43 ± 0.58	1.40 ± 0.61
Mean dominant frequency (kHz)	1.28 ± 0.21	1.12 ± 0.15
Dominant frequency range (kHz)	11.68 ± 4.61	4.86 ± 2.45
Modulation index (n/a)	72.11 ± 32.59	56.98 ± 32.21
Call duration (s)	22.31 ± 16.15	4.65 ± 1.27
Inter-quartile range (time; s)	10.32 ± 8.35	1.82 ± 0.61
Avg. # of syllables (n/a)	72.18 ± 59.97	6.18 ± 4.46
Spectral skew (n/a)	3.99 ± 0.67	4.68 ± 1.32
Shannon entropy (0–1 scale)	0.91 ± 0.01	0.94 ± 0.02

Petter & Charles-Dominique, 1979; Turner & Harrenstien, 1985). Because of the strong fission-fusion dynamics exhibited by ruffed lemurs, roar-shrieks appear to maintain social cohesion within a community that is split up into different subgroups without visual contact (Batist et al., 2022; Macedonia & Taylor, 1985; Petter & Charles-Dominique, 1979; Turner & Harrenstien, 1985). Barks have been previously described as alarm calls for aerial or terrestrial predators (Macedonia, 1990; Pereira et al., 1988; Petter & Charles-Dominique, 1979; Turner & Harrenstien, 1985).

Our goal for this study was to establish a comprehensive PAM workflow for ruffed lemur loud calls in southeastern Madagascar, that could also serve as a guide for other primatologists. And beyond data collection, we also aimed to develop a deep learning analytical pipeline. PAM has been transformed by the “big data” revolution, recent advances in artificial intelligence, and development of highly efficient and potent machine learning (ML) models (Dufourq et al., 2022; Stowell, 2022; Tuia et al., 2022). Deep learning models such as convolutional neural networks (CNNs), have proven to be useful in creating passive acoustic monitoring classifiers (Stowell, 2022), and have successfully been used in diverse species including primates (Table 1; Dufourq et al., 2021; Ruan et al., 2022), bats (Chen et al., 2020; Mac Aodha et al., 2018; Roemer et al., 2021), whales (Bermant et al., 2019), and birds (Grill & Schlüter, 2017; Gupta et al., 2021; Kahl et al., 2021; Lauha et al., 2022). In these models, a computer “trains” itself to recognize species-specific calls by extracting salient features from a reference data set (here, spectrograms) and then classifies segments from other test recordings based on their similarity to those features. This innovative automated approach allows thousands of hours of recordings to be analyzed efficiently rather than via labor-intensive manual processing (Stowell, 2022; Tuia et al., 2022). The ruffed lemur workflow described herein is among the first studies to use an integrated PAM and ML pipeline with any lemur species (see also Ravaglia et al., 2023). Our study had three main objectives. First, we aimed to



**FIGURE 1** Spectrograms of black-and-white ruffed lemur bark (a) and roar-shriek (b) from active acoustic monitoring (directional microphone) in Mangevo, Ranomafana National Park, Madagascar.

develop a machine learning pipeline for automated detection of ruffed lemur roar-shrieks, as a proof-of-concept for future scaled-up projects. Second, we wanted to compare roar-shriek detections from PAM and from in-person observations (focal follows) to demonstrate its effectiveness. To further emphasize the value in PAM data, our final aim was to assess diel patterns in ruffed lemur calling behavior and determine if ruffed lemurs exhibited nocturnal vocal activity (as PAM devices ran continuously). Based on previous anecdotal accounts, we predicted that ruffed lemurs would call at night, but to a lesser extent than during the day.

## 2 | MATERIALS AND METHODS

### 2.1 | Ethical note

The research herein adheres to the American Society of Primatologists' *Principles for the Ethical Treatment of Primates* and was approved by the Hunger College Institutional Animal Care and Use Committee

(IACUC; protocol no. AB-2/22). This project was approved and permitted by Madagascar National Parks and the Malagasy Directorate of Protected Areas and Ecosystems (permit no. 109/19/MEDD/SG/DGF/DSAP/SCB.Re).

#### 2.1.1 | Data collection

We conducted fieldwork at Mangevo (21.3833S, 47.4667E), a remote, pristine-forest site in Ranomafana National Park (RNP), southeastern Madagascar, from May–July 2019 (Supporting Information S1: Figure 1). Mangevo is characterized as a mid-elevation tropical rainforest (800–1100 m; Wright et al., 2012). Mangevo is the site of the long-term Ranomafana Ruffed Lemur Project (Baden et al., 2016). There is a habituated ruffed lemur community at Mangevo that, at the time of the study, contained 31 radio-collared individuals (for details on collaring see Glander, 1993) during the majority of the study (two were killed by fossa, *Cryptoprocta ferox*, in July).

For passive acoustic monitoring, we used four autonomous recording units (ARUs): two SongMeter SM4s (hereafter SM; Wildlife Acoustics) and two Swifts (Cornell Yang Center for Conservation Bioacoustics). Recorders were placed within the core areas of known subgroups, with a minimum distance of 300 m between each device (Supporting Information S1: Figure 2). We opted for this distance between each recorder for a number of reasons: (1) Mangevo is a dense, montane rainforest so sound attenuation was expected to be very high due to topography and vegetation; (2) roar-shrieks are highly variable (in bandwidth, duration, amplitude) and amplitude is affected by how many lemurs are calling simultaneously (can be 2–7 individuals; Batist et al., 2022); and (3) to account for variation in microphone sensitivity across recorder types. Placing the recorders relatively close to one another was beneficial as it ensured that many calls were recorded despite such variability in call amplitude and recorder sensitivity (and while still removing potential pseudo-replicates; see below for further detail). The Swifts were built into a Pelican case (the “rugged” version), and the SMs were placed into the Wildlife Acoustics metal housing. Each recorder was placed as high in a tree as was safely possible (~2–3 m), microphone oriented sidewise, and secured with cable locks. This is high enough to avoid any ground interference, but still allows to arboreal species' calls because the microphone is omnidirectional (in both horizontal and vertical space). We used NiMH rechargeable “D” batteries (brand: EBL); the Swifts can hold nine D batteries while the SMs can hold four. We used SanDisk ExtremePro SD cards in both devices. The SM and Swift recorders had sampling rates of 48 and 32 kHz respectively and recorded in 16-bit mono format. The microphone sensitivity for SMs is –35 and –44 dB for Swifts (taken from manufacturer spec sheets). We used a gain of +16 dB for SMs and +20 dB for Swifts. We used an increased gain for the Swift to better standardize the recording specifications given the higher sensitivity of the SM. The different devices did not have matching sample rate settings (due to differences in available settings between Swifts and SMs), but we downsampled all recordings in a preprocessing step to standardize them (see below). ARUs were scheduled to run continuously, 24 h a day. We checked each device once every 2–3 weeks to ensure proper functionality and change out batteries and SD cards when necessary. Due to the remoteness of the sites and requirements of the focal follow sampling (see below), we were not able to check all devices on a standardized schedule. In some cases, devices had stopped recording due to battery or SD card failure before we could replace them in time. In total across all recorders, we collected approximately 2300 h of recordings across 55 days.

Concurrently with the passive acoustic monitoring, we also conducted 50 full-day focal follows (Altmann, 1974) of radio-collared individuals, located daily via radiotelemetry (375 observation hours). The exact start time of each focal was variable depending on how long it took to locate the focal individual at the beginning of each sampling period. We targeted each individual as a focal animal at least once (and maximum 3x) over the study period to avoid sampling bias. The focal was never the same individual on consecutive days. We recorded all occurrences (Altmann, 1974) of ruffed lemur

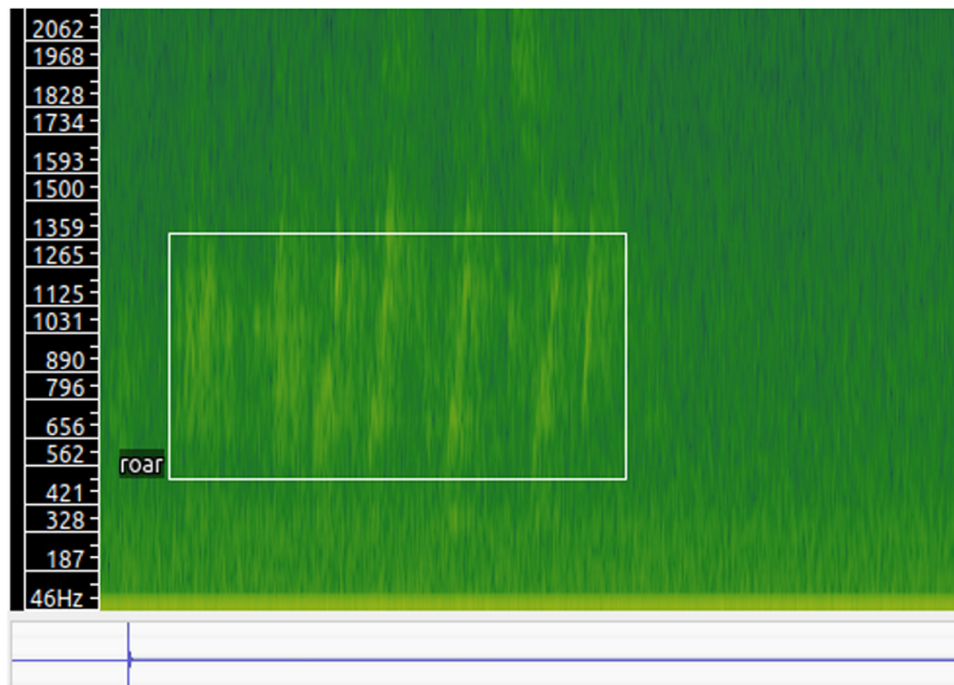
vocalizations (i.e., not just those given by the focal animal), including calls heard from out-of-sight subgroups. Where possible, we recorded a GPS point and caller(s) demographic information (e.g., individual ID, sex), for each vocalization (sometimes not possible if individuals were not visible or on the move). As mentioned previously, we focused on only roar-shrieks and barks (*sensu* Batist et al., 2022) for this specific study comparing passive acoustic monitoring and in-person observations.

### 2.1.2 | Data analysis—Annotation of the acoustic recordings

Initially, we had 2300 h of unlabeled audio data recorded over multiple files, where the files were either 20 (SM) or 40–60 (Swift) min in duration. We began by randomly selecting audio files to manually annotate using Sonic Visualiser (Cannam et al., 2010) via visual inspection of the spectrograms and simultaneously listening to the audio. We annotated acoustic events which had ruffed lemur roar shrieks as presence events, and other sounds as absence events. The manual annotations were conducted independently by the three authors: C. B., E. D., and L. J. Subsequently, a collaborative verification was performed to ensure consensus among the three annotators, resulting in the final annotations. To annotate a lemur roar, labeled as presence, or a soundscape with or without other sounds, labeled as absence, we drew bounding boxes around the specific event as depicted in Figure 2, allowing us to record the start and end time of each window.

While accurate annotation of the start and end times was imperative, we did not retain the lower and upper frequency parameters of the bounding boxes due to the utilization of a predefined frequency range (0.5–4 kHz). Our findings and prior research (Dufourq et al., 2021, 2022; Jeantet & Dufourq, 2023; Kahl et al., 2021; Ruan et al., 2022) revealed that it was not necessary to include the full spectrogram, and that improved performance was obtained when narrowing the range. Specifically, our knowledge that ruffed lemurs vocalize within the 1–2 kHz range prompted us to define this range between 0.5 and 4 kHz. This decision aimed to guide the model's attention to the frequency range containing lemur vocalizations, allowing it to discern distinctive features of lemur calls. Additionally, this range was chosen to facilitate the model's learning about other vocalizations that might overlap with lemur sounds (which explains why we extended the range slightly beyond 2 kHz). Furthermore, we also noticed the presence of harmonics extending up to 4 kHz.

To train the model to identify the lemur calls and distinguish them from the other sound present in the soundscape, it is crucial to train the model to identify biophony, geophony, and anthropophony sounds, in addition to ruffed lemur calls. However, using the entire absence sequence of the audio files to create this absence class would generate a significant amount of data and increase the processing time to train the detection model. Although randomly selecting a fixed number of absence windows is an option, our research has shown that this approach would increase the number of



**FIGURE 2** Example of a bounding box drawn around a ruffed lemur roar-shriek in Sonic Visualiser.

false positives due to the low diversity of selected sounds from random selection. Most of the time, the audio files contain soundscape sounds without any particular sound of interest. Sounds of interest that overlap with the range of lemur shrieks are more spontaneous, and therefore more difficult to randomly select. By manually identifying other sounds, we create a more diverse set of examples for the absence class which enables the algorithm to become more robust to false positives. Thus, to ensure that relevant sounds are not missed and to keep the data set computationally feasible, we manually annotated windows containing these sounds (as detailed in Dufourq et al., 2021).

### 2.1.3 | Data analysis—Preprocessing of acoustic recordings

To standardize the sampling rates (32 kHz for the Swifts and 48 kHz for the SM), we first applied a low-pass filter to each recording with a cutoff rate of 4 kHz and subsampled them to obtain a new sampling rate of 9.6 kHz. Based on the manual annotations, we extracted labeled windows and divided them into audio segments through a sliding window of 4 s length with an overlap of 1 s. The segments were then converted into mel-frequency spectrograms using a Hann analysis window size of 107 ms (1024 samples) with a hop size of 27 ms (256 samples) and 128 mel-frequency. This resulted in spectrogram images of size  $128 \times 151$ . We thus created a binary classification data set (presence and absence of lemur vocalization).

### 2.1.4 | Data analysis—Deep learning model

Deep learning refers to algorithms, commonly called deep neural networks, which possess the capability to autonomously identify complex and highly distinctive patterns within data. Deep learning algorithms are sequences of processing layers that perform a combination of linear and nonlinear mathematical operations. These interconnected layers encompass multiple model weights, represented as real-valued numbers (such as  $-0.2$  or  $1.7$ ), which undergo optimization throughout the training phase of the model. More precisely, for a classification problem, the model is first generated with random initial weights and then provided with a set of annotated data during its training phase. During this training phase, the weights are iteratively adjusted to optimize the network's ability to generate accurate target outputs for a given input. By increasing the number of layers, the network is able to learn an abstract representation between the input and output. The layers closer to the input would be referred to as “earlier layers” and those closer to the output as “later layers” which learn a much more abstract representation of the input due to the chained operations of consecutive layers. Deep neural networks applied to image classification tasks are typically referred to as convolutional neural networks (CNNs). These networks contain convolutional layers which have the ability to learn visual features from the input.

A nuanced aspect of deep learning emerges from the observation that as a model grows in complexity—defined by an increasing number of layers and consequently weights—it demands a more extensive training data set. An undersized training data set frequently

results in overfitting issues, wherein the model performs well on the training data but poorly with new data, due to becoming excessively tailored to the training data's specifics. Hence, the use of available robust and large size deep learning algorithms, is often limited in ecology by the scarcity of labeled data. But recently, transfer learning has shown promising results in mitigating this challenge in bioacoustics (Bravo Sanchez et al., 2021; Dufourq et al., 2022; Zhong et al., 2020).

Within the context of CNNs, image and audio classification, transfer learning consists of using an existing model (say model *M*), that has already been trained to perform a task (say task *A*) similar to the one expected (say task *B*). Generally, task *A* will have a corresponding larger data set than task *B* (Dufourq et al., 2022; Zhong et al., 2020). When implementing transfer learning, one obtains the weights from the existing model *M*, and these weights are used as the starting point (as opposed to randomly initialized weights) before the training of task *B*. At this stage, the model is modified so that it can be adapted to task *B*. The earlier layers are usually kept as is, and they are "frozen." This implies that the frozen weights in earlier layers are not updated during training. These layers are sometimes referred to as the feature extractor, as these layers have already been trained to learn visual features from images. The weights in later layers of *M* are not frozen, and thus those weights can be optimized. The process of updating the weights in the later layers is known as "fine-tuning." These later layers are known for acquiring the skill to leverage the visual features for classification in a particular context. Through transfer learning, we harness the model's overarching capability to extract input information and exclusively fine-tune the final layers to adapt it to our specific scenario. In this study we used transfer learning since we did not have a large number of annotated lemur examples, and thus, task *B* would be a binary classification problem for lemur vocalizations.

The ResNet152V2 architecture (He et al., 2016) is a well-known CNN which has successfully been used in various application areas. Typically, this network is used within the context of transfer learning and has been pre-trained on ImageNet, a database comprising over 14 million human-annotated images (Deng et al., 2009). This architecture has produced good classification performance on various PAM classification tasks (Bergler et al., 2022; Dufourq et al., 2022). More specifically, we downloaded the pretrained ResNet152V2 ImageNet weights and froze the feature extractor (all convolution and pooling layers which are used to extract visual features from the input). The output of the convolutional layers was flattened and we added a fully connected layer with 2 units and a softmax activation. Softmax activation normalizes the output of a network into the probability that the spectrogram belongs to each class.

The final class (presence or absence) was assigned to the class with the highest probability. The weights in the fully connected layer were randomly initialized. There were thus 58,331,648 nontrainable weights in the feature extractor, and 81,922 trainable weights in the fully connected layer. This CNN expects a three-channel input image, and thus each single channel spectrogram was manipulated into three channels in the same manner described in Dufourq et al. (2022) by

creating two additional normalized spectrograms using exponents of the original spectrograms. Figure 3 provides an overview of our approach.

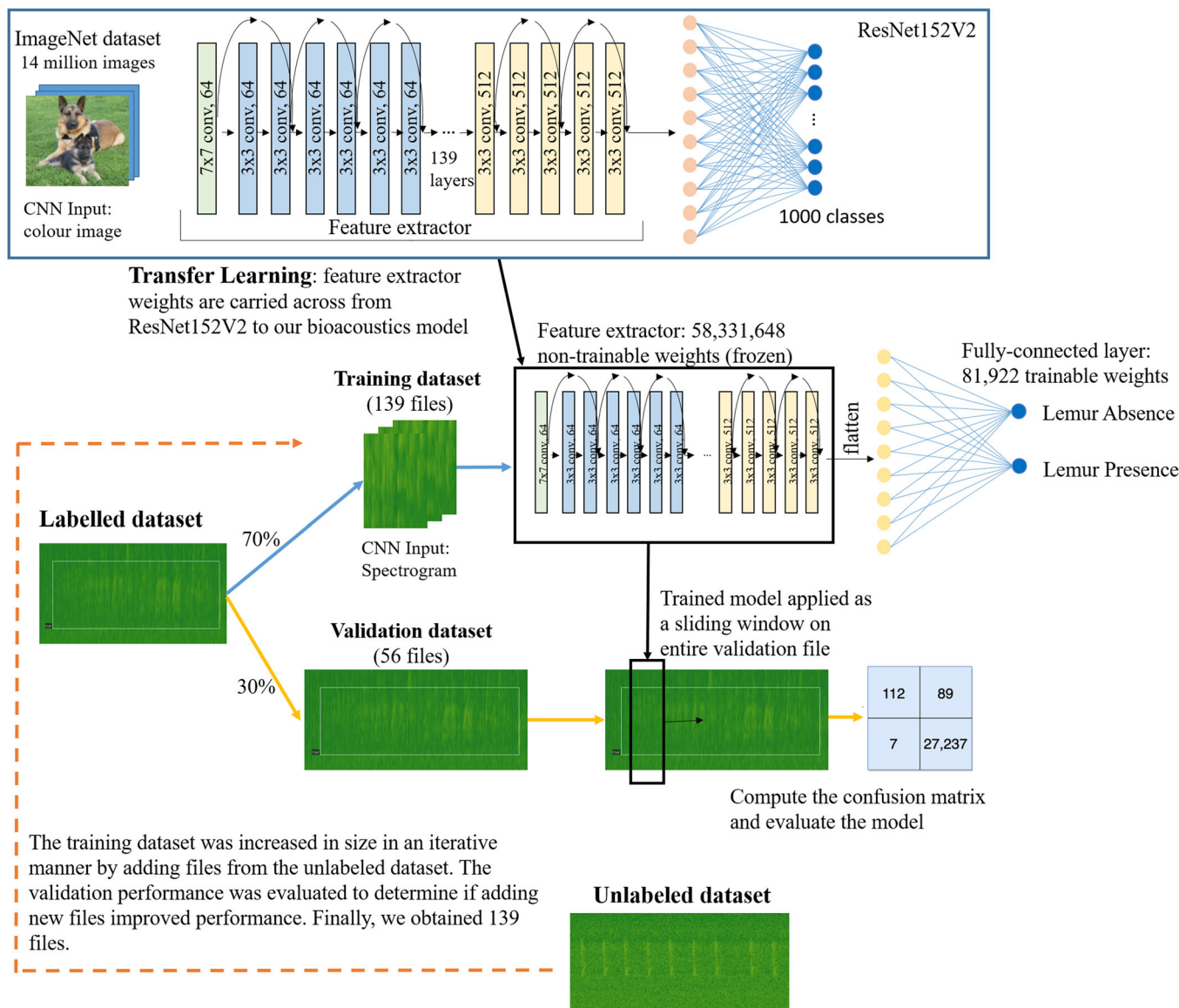
### 2.1.5 | Data analysis—Training and evaluation

We randomly split the labeled audio files into training (70%) and validation (30%) data sets. Entire audio files were assigned to either the training or validation set, that is, we did not assign a portion of an audio file to training and another portion to validation. Our preliminary findings revealed that we obtained various false positives using the model trained on the initial training data set. Therefore, we iteratively augmented the training data set by incorporating additional files from the unlabeled set until the validation performance converged and no longer improved. To include these unlabeled files, we applied the CNN to approximately 10 of them (the number of files added in each iteration varied slightly), and manually validated the predictions. Once validated, these files were added to the training data set. The false positives, mostly corresponding to biophony and anthropophony, were manually labeled as the absence class, while the identified lemur calls were labeled as the presence class. Subsequently, the CNN was retrained using these newly annotated examples. This was repeated three times until the performance on the validation set was no longer improved. In the end, 195 files (114 h) were annotated, with a training data set of 139 files (approximately 84 h, 51 files from the SM recorder and 88 from the Swift recorders), and a validation data set of 56 files (approximately 30 h, 26 files from the SM recorder and 30 from the Swift recorders). We obtained 4856 presence and 6979 absence windows for the training data set.

The validation data set was created from 56 audio files (length 20, 40, or 60 min) and had 119 presence events where lemur calls were manually detected. The remainder of these audio files consisted of absence events (approximately 30 h of absence corresponding to 27,326 absence windows). We do not consider absence windows for the validation data as we applied our model to each of the 56 validation audio files as though this were a real-world case-study. To determine the overall model performance, and similarly to a real-word case study, we applied the model on the entirety of each validation file using a sliding window. For each file, we started at the beginning, and extracted a 4 second window (seconds 0–4). We created a spectrogram (see above), and the model was then applied to that spectrogram where a prediction was generated. We then slid to seconds 1–5, created a spectrogram, and predicted again. This was repeated until the end of the file was reached.

We then postprocessed the files (as detailed in Dufourq et al., 2021) to obtain the start and end times for each detected calling bout. We compared predicted calling bouts to the annotated calls and calculated a confusion matrix. Based on the confusion matrix, we calculated the precision, recall, and accuracy of the model. There was an inherent trade-off between precision (of all presence predictions, how many were actually presence) and recall (of all presence events, how many were actually detected). We developed the model with a focus on optimizing for a high recall, as the threat of potentially missing a call (and





**FIGURE 3** Illustrating how transfer learning was used to classify between lemur and nonlemur vocalizations. We adapted ResNet152V2 by freezing the feature extractor, and then adding a fully connected layer with 2 softmax units (presence or absence). The unlabeled data set was used to incrementally supplement the training data set until we obtained 139 training files. The model was trained on the 139 files, and evaluated on 56 audio files.

stating that site lacks ruffed lemurs) was higher than the expense of more false positives to manually process.

### 2.1.6 | Data analysis—Application on the unlabeled data set

To evaluate the performance of the CNN on the unlabeled data set, we applied the model on the entirety of each unlabeled file using the same approach described above (2300 h, ~900 GB). All the predictions were manually verified. The CNN was trained on an Nvidia RTX A4000 GPU (8GB RAM) and was implemented in Tensorflow 2 (Abadi, 2016) and Python 3. Audio processing was done using Librosa (McFee et al., 2015) and was executed on a computer running Ubuntu 22.04.2 with 64GB

RAM. We used the Adam optimizer, the categorical cross entropy loss function, a batch size of 32, and a learning rate of 0.000001. The CNN was trained for 5 epochs and took 50 s to train. These hyper-parameters were obtained via preliminary experiments and by considering previous research that used transfer learning (Dufourq et al., 2022). The same hardware was used when we applied the model on the unlabeled data set.

### 2.1.7 | Data analysis—Diel patterns and methods comparisons

Upon completing our CNN pipeline, we qualitatively compared passive (PAM) and active (in-person observations) acoustic

**TABLE 3** Confusion matrix for the validation data set.

		Actual	
		Presence	Absence
Predicted	Presence	112	89
	Absence	7	27,237

monitoring with respect to the number of ruffed lemur calls detected, total cost, temporal coverage, spatial coverage, and other miscellaneous requirements in the context of this study. We assessed the number of calls across the days, and additionally, we compared the number of calls across hours of the day to assess any diel patterns in ruffed lemur vocal activity. To remove any possible pseudo-replications (the same call picked up on different recorders), we filtered out calls that occurred less than 3 s apart at multiple recorders ( $n = 670$ ). The study period encompassed the ruffed lemur mating period (only a few days/year as ruffed lemurs, like most lemur species, are highly seasonal breeders; Rasmussen, 1985), so we were particularly interested in assessing vocal activity rates during this time. Because of the low sample size (only three recorders, short sample period), we just report descriptive statistics for vocal activity and diel patterns.

### 3 | RESULTS

One of the SMs failed early on in the study, so the data collected (and presented here) are from the remaining three ARUs (SM, Swift1 [S1], Swift2 [S2]). On the 56 validation audio files (roughly 38 h, 8.7 GB), the best CNN model obtained an accuracy of 90.0%, resulted in 89 false positives (roughly 6 min), 7 false negatives and 112 true positives (Table 3 provides a confusion matrix). The model achieved an F1-score of 0.7 on the validation set. Thus, on 38 h of data, approximately 6 min would have been wasted on verifying false positives. It took approximately 18 min to predict on all validation files which includes preprocessing of each file and applying the model on the windowed spectrograms. On the validation data set, the model obtained a precision and recall of 55.7% and 94.1% respectively.

We then used this model on the entire unlabeled data set, and manually validated the model predictions. Using an Nvidia RTX A4000 and Intel CPU (i7-11850H) to predict on the entire data set took 27 h for the SM data, 15 h for S1 and 18 h for S2 (60 h total). Access to high-end GPUs is not always feasible, and thus we also explored the computational effort using just a CPU. A single Swift audio file (60 min) took 149 s to process and an SM file (20 min) took 59 s. Therefore, it would have taken about 150 h in total to predict on the entire data set using a CPU.

After manual validations of the model predictions to determine false positive and true positive rates, we calculated the precision. Precision was 87.9% for S1, 91.0% for S2, and 23.0% for SM. We expected lower precision for SM as it was closer to a river and (unbeknownst to us) where our campsite leads chopped wood

(further details given in Section 4.1 under Section 4). Across all recorders, rain was the incorrect sound type most predicted as a lemur (46.0% of all false positives) followed by wood chopping (25.0%) and bird calls (13.0%). Rain was the most-confused sound for SM and S1, while bird calls were the most-confused for S2.

To foster accessibility, knowledge-sharing, and user-friendly tools, we developed an open-source graphic user interface (GUI) which facilitates the use of the model to detect lemur vocalizations (Supporting Information S1: Figure 3). While this model has been trained on lemurs, it could be generalized to and trained on other calling species; similarly, the GUI is flexible and could be re-used with different models.

#### 3.1 | PAM versus in-person observation

PAM resulted in more detected barks and roar-shrieks (absolutely and per unit time) and more overall sampling hours than in-person observations (Table 4; Figure 4). PAM was also more cost-effective and required less time (12 vs. 56 days) and labor to implement (Table 4). The spatial scale of both methods was fairly limited due to ruffed lemurs' small home range sizes, and PAM was further affected by the malfunction of one SM (Baden et al., 2020). But, in-person observations will inherently offer greater flexibility in spatial scale as PAM is restricted to stationary recorders. Beyond data collection, using the deep learning pipeline significantly decreased the time needed for data analysis as well (11 weeks for manual processing vs. 2 weeks for deep learning; Table 4).

#### 3.2 | Diel patterns

There does not appear to be a diurnal pattern in vocal activity; unlike many other primate species, ruffed lemurs in this study did not exhibit a peak in calling activity at dawn and/or dusk. Roar-shrieks and barks were given throughout the day at a roughly-constant rate from 7:00–17:00, though roar-shrieks appear to occur more frequently later versus earlier in the day (Figure 5). We found that ruffed lemurs do exhibit nocturnal vocal activity, albeit still at a far lower rate than during the day (Figure 5). Upon further inspection, we found that nearly all nocturnal vocal activity coincides with the ruffed lemur mating period (Figure 6). We observed mating on June 29, July 3, July 8–11 and July 13; this generally aligns with peaks in both diurnal and nocturnal calling rates. However, because of ruffed lemurs' fission-fusion dynamics (individuals within a community move between subgroups), we are not able to tie particular calls to a specific individual or subgroup. Unfortunately, only SM was running in the second half of July (postmating season), and as previously discussed, had a smaller detection range due to higher background noise. Nevertheless, the SM call detections from the third week of July show no nocturnal calls and a small number of diurnal calls, mimicking the pattern observed before the mating period. Given that this is only from one recorder, these results are very preliminary and should be interpreted with caution.

**TABLE 4** Comparison in scale and scope of sampling design between passive acoustic monitoring and on-the-ground observations.

	Passive acoustic monitoring and machine learning	On-the-ground observations and manual processing
Total # of call (bark, roar-shriek) detections	5118	740
Call detections per unit time (h)	2.23 (5118 detections/2300 h)	1.85 (740 detections/400 h)
Cost (USD)	\$720–\$1820 <ul style="list-style-type: none"> <li>Recorders x3—\$300–\$1200 (with Audiomoths, would be \$300)</li> <li>Batteries (NiMH “D”)—\$100–\$200 (cost variable between brand, type (AA vs. D) alkaline vs. NiMH)<sup>a</sup></li> <li>SD cards—\$50–\$150 (cost variable between size, brand, microSD vs. SD)</li> <li>Field team costs—\$270</li> </ul>	\$1950 <ul style="list-style-type: none"> <li>Field team salaries—\$1250</li> <li>Food—\$500</li> <li>Supplies—\$200</li> </ul>
Time (labor hours)-data collection	12 days (optimal # if only doing PAM) <ul style="list-style-type: none"> <li>4 trips—Deploy recorders, check 1 [batteries/SD], check 2, retrieve)</li> <li>3 days/trip—Hike in, recorder work, hike out</li> </ul>	56 days <ul style="list-style-type: none"> <li>50 full-day focal follows (x ~ 8.5 h per) = 400 h</li> <li>6 days—Hike to/from site</li> </ul>
Time (labor hours)-data analysis	~2 weeks <ul style="list-style-type: none"> <li>Annotating training data-7 days</li> <li>Running model-4 days (CPU)</li> <li>Validating predictions-3 days</li> </ul>	~11 weeks <ul style="list-style-type: none"> <li>2300 h raw data × 15 min/h<sup>b</sup> = 34,500 min (575 h) to manually validate entire data set</li> <li>575 h = 11 weeks (assuming 50-h work week)</li> </ul>
Temporal sampling scale	<ul style="list-style-type: none"> <li>2300 recording hours across 2.5 months</li> <li>24-h coverage</li> </ul>	<ul style="list-style-type: none"> <li>400 direct observation hours across 2.5 months</li> <li>12-h coverage</li> </ul>
Spatial sampling scale	~1 km <sup>2c</sup>	~1.5 km <sup>2d</sup>
Miscellaneous	Other benefits <ul style="list-style-type: none"> <li>Permanent record that can be re-analyzed for other use cases, or assessing change over time</li> <li>Record all other vocal species too</li> </ul> Other considerations <ul style="list-style-type: none"> <li>Only loud/long calls used for ML detection<sup>e</sup></li> <li>Less specific data</li> <li>Tech malfunctions sometimes</li> </ul>	Other benefits <ul style="list-style-type: none"> <li>All call types recorded</li> <li>More contextual data (group size/composition, behavior, etc.)</li> </ul> Other considerations <ul style="list-style-type: none"> <li>(Mangevo)—Observations facilitated because of habituated, radio-collared groups; requires VHF receivers/antennas (\$1000 s of dollars)               <ul style="list-style-type: none"> <li>Projects without this set-up would require even more manual effort and time</li> </ul> </li> </ul>

<sup>a</sup>Alkaline batteries are cheaper (by ~50%) than NiMH batteries, but cannot be reused and increase electronic waste. So, this cost is variable and could be lower if using alkaline.

<sup>b</sup>To an experienced listener (CHB), a combination of visually and aurally inspecting spectrograms manually took ~15 min per hour of recordings. A naïve listener would likely take longer, further elongating the time required for data processing.

<sup>c</sup>Approximate detection radius (Swifts-350 m, SM-150 m).

<sup>d</sup>Home range sizes of Mangevo groups, from Baden et al. (2020).

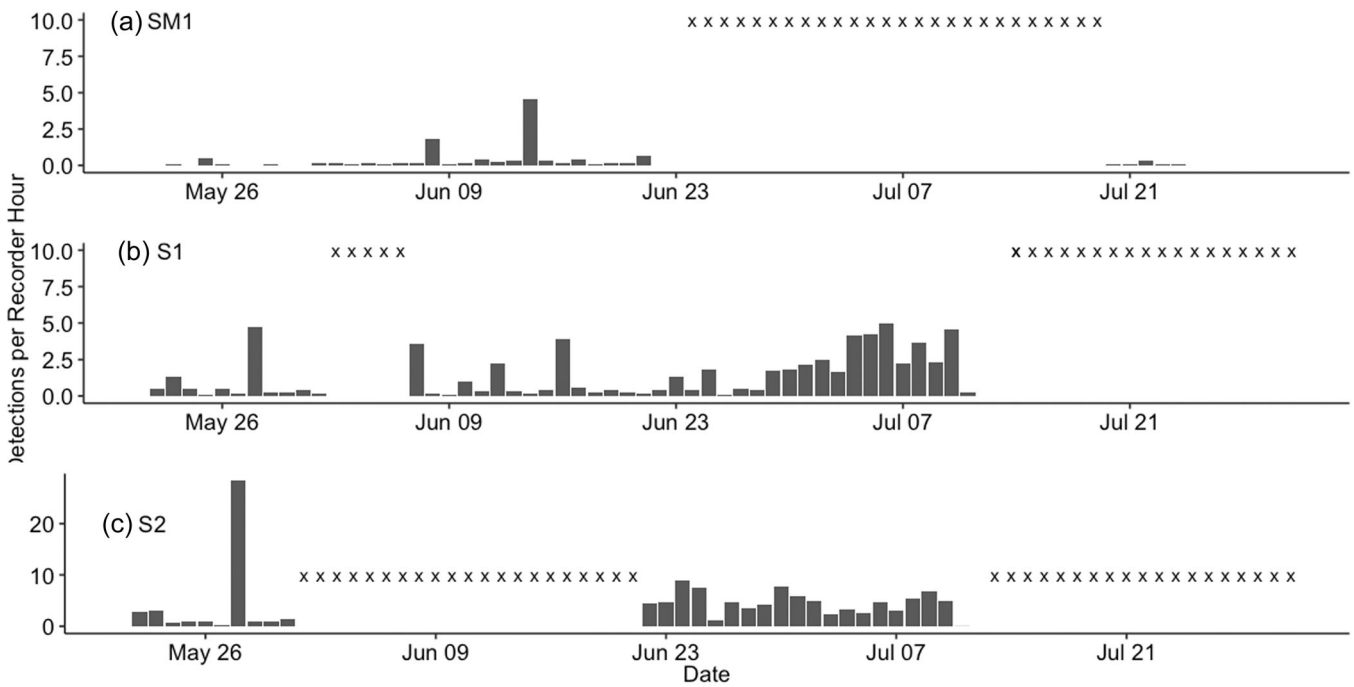
<sup>e</sup>PAM recorders could have detected other ruffed lemur calls but we focused only on the two loud calls in our ML pipeline.

## 4 | DISCUSSION

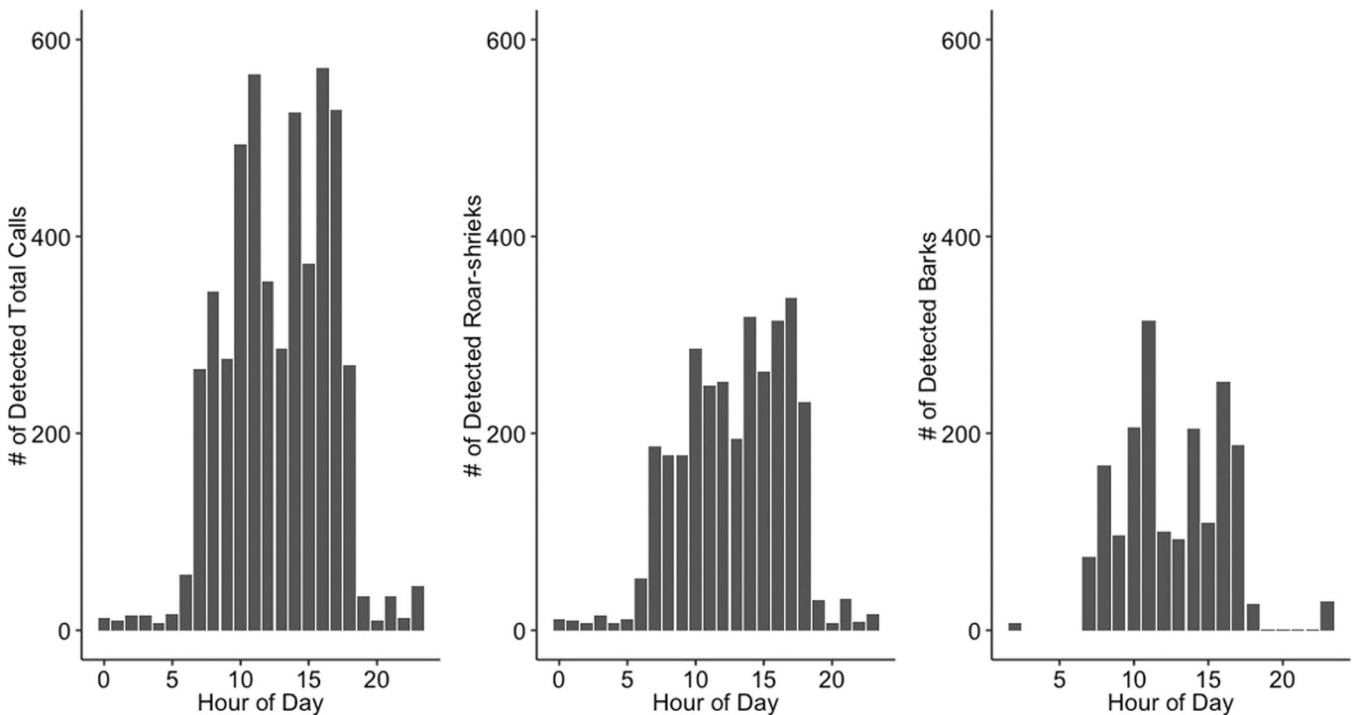
In this study, we combined powerful new technical advances such as passive acoustic monitoring and deep learning to study black-and-white ruffed lemurs in southeastern Madagascar. Our results suggest that PAM is a powerful and cost-efficient method for detecting ruffed lemurs in their natural habitat. This is in line with previous studies that have shown PAM can be as accurate and efficient as traditional monitoring methods, if not more so (Boullhesen et al., 2021; Castro et al., 2019; Darras et al., 2019; Digby et al., 2013; Leach et al., 2016; Melo et al., 2021). We found PAM to be more effective than in-person observations with respect to time, cost, labor, and the

benefit of yielding a data set that can be re-analyzed for other use cases. While PAM requires a high upfront cost to purchase autonomous recorders, batteries, and SD cards, this paid for itself after only 2 months in the context of our study. However, this study was conducted at one field site with a small survey area so this may not generalize to larger, landscape-scale projects. We do acknowledge that the high upfront cost for such large projects can still be prohibitive, but the cost of in-person observations for such a project would likely be just as high, if not higher.

The in-person observations for this project are also facilitated by the fact that the Mangevo ruffed lemur population is habituated, radio-collared and therefore easier to find and follow. Ruffed lemurs also have



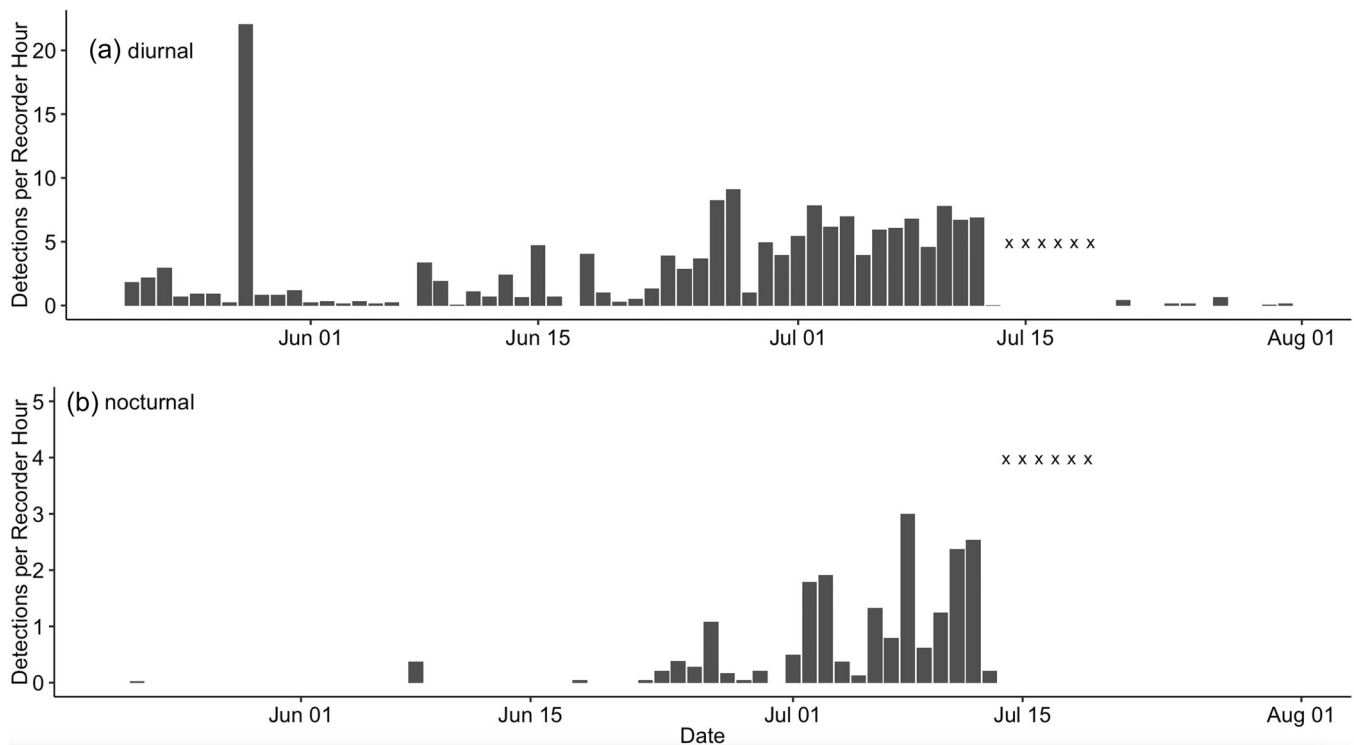
**FIGURE 4** Vocal rate of black-and-white ruffed lemur calls (roar-shrieks + barks) per day, per recorder in Mangevo, Ranomfana National Park. (a) is Swift1 recorder, (b) is Swift2 recorder, (c) is SongMeter1 recorder. Vocal rate = detections/recorder hours, where recorder hours is the total number of hours each recorder was running.



**FIGURE 5** Total number of black-and-white ruffed lemur calls (left panel), roar-shrieks (middle), and barks (right) per hour in Mangevo, Ranomafana National Park.

fairly small home ranges (and core areas within those; Baden et al., 2020), so more effort would be required for more migratory species, regardless of monitoring method. One of the greatest benefits of PAM, though, is the permanent record of the soundscape, which allows the data set to

be analyzed for different species and use cases in future. This audio data can also serve as an important baseline with which to compare changes at the same site over time. Of course, PAM misses out on critical contextual and behavioral information that in-person observation allows



**FIGURE 6** Vocal rate of black-and-white ruffed lemur calls (roar-shrieks + barks) in Mangevo, Ranomafana National Park, per day, split between (a) diurnal (6:00–18:00) and (b) nocturnal hours (18:00–6:00). Vocal rate = detections/recorder hours, where recorder hours is the total number of hours each recorder was running.

for (Batist et al., 2022; Pereira et al., 1988). It is this information that permits primatologists to infer call function. PAM is also biased towards a species' long calls, and misses much of the softer calls within the repertoire. Taken together, we consider active and passive acoustic monitoring to be complementary techniques that can each have strengths and weaknesses depending on the research question or use case.

As primates are quite soniferous on the whole, PAM presents an extremely promising opportunity for primatologists in particular. The fragmented nature of the many primate species' habitats makes traditional monitoring methods challenging, and PAM could provide a solution to this problem. By using PAM, primatologists can obtain long-term, noninvasive data for multiple primate species in an area simultaneously. Such presence/absence data, in combination with geospatial variables, could then be used to generate occupancy and species distribution models (Appel et al., 2023; Campos-Cerqueira & Aide, 2016; Kalan et al., 2015). PAM can also yield important insights beyond simple presence metrics; for example, without PAM we would not have quantitative evidence of or patterns in ruffed lemur nocturnal vocal activity. Furthermore, our results indicate that ruffed lemurs do not exhibit a dawn/dusk chorus, which is in line with previous research demonstrating no diel patterning to ruffed lemur diurnal vocal activity (Batist et al., 2022; Geissmann & Mutschler, 2006). Our present study expands on this by providing quantitative evidence of nocturnal calling behavior, the majority of which occurred during the mating period.

Our study also highlights the potential of deep learning models such as CNNs in creating passive acoustic monitoring classifiers (Stowell, 2022; Tuia et al., 2022). We trained and implemented an open-source CNN classifier using transfer learning for detecting ruffed lemur calls. This demonstrates the potential of using machine learning models to automate the detection and classification of species-specific calls in PAM data, reducing manual effort and improving the efficiency of data processing and validation (Dufourq et al., 2022; Ruan et al., 2022; Stowell, 2022). Transfer learning, in particular, holds great potential for bioacoustic classification tasks which are typically characterized by class imbalance, long tails, and small training data sets (Bravo Sanchez et al., 2021; Zhong et al., 2020). As previously mentioned, we optimized the CNN to maximize recall, so that we had a higher probability of detecting as many calls as possible. We achieved an F1-score of 0.7 and a remarkably high recall of 0.94 even with a small (by machine learning standards) training data set. Training data are often difficult to obtain in bioacoustics, particularly for endangered or elusive species. Transfer learning, and other even more recent advances such as autoencoders and transformer models, can help overcome this common obstacle for PAM analysis (Dufourq et al., 2022).

While the application of deep learning in PAM research has gained traction, it is crucial to emphasize the importance of accessibility and the facilitation of knowledge sharing within the scientific community. We should promote the development of user-friendly tools that go beyond the mere release of code on

platforms such as GitHub. While sharing code repositories is valuable for collaboration and reproducibility, it can present barriers for noncoders who may struggle with the technical intricacies of implementation. To overcome this, it is essential to encourage the creation of comprehensive tools and frameworks that offer intuitive interfaces, detailed documentation, demos (via e.g., Google Colab), and step-by-step user guides. We should actively encourage the Python (and scientific research at large) community to develop open-source and customizable GUIs. These GUIs would empower researchers, regardless of their coding expertise, to harness the power of deep learning algorithms in bioacoustics analysis and primate research. Existing GUIs which use machine learning include Deep Audio Segmenter (Steinfath et al., 2021) which allows a user to annotate and train a deep neural network, and BirdNET-Analyzer (Kahl et al., 2021) which is integrated into Raven Pro (K. Lisa Yang Center, 2023). By promoting open-source and accessible initiatives, we can create a collaborative space where researchers, irrespective of their programming capabilities, are empowered to contribute to the refinement and improvement of PAM methodologies. We can empower a broader audience of researchers to leverage deep learning approaches effectively in the context of PAM and primate studies.

#### 4.1 | Limitations and future directions

Our study was limited to three recording devices in one small survey area, so the results (particularly related to diel patterns) should be taken with caution. Due to these sample size limitations, we were not able to statistically test vocal activity patterns and only report qualitative results. The vocal activity patterns during the short mating period are fairly anecdotal in that (unfortunately) only one recorder was functional in the postmating period. However, given the extremely short mating period, it is very difficult to collect data during this time. We hope that our preliminary results will spark future research in ruffed lemur mating behavior and communication.

Additionally, using different recorders introduced inherent variation based on differing technical specifications, and we would recommend researchers use one type of recorder for future studies to mitigate this. However, the overall pipeline described here is broadly scalable to larger areas and to other primate species. Indeed, a broader-scale PAM study (using this paper's workflow) across the southeastern Malagasy rainforest corridor is currently in progress to maximize the scope of ruffed lemur monitoring in the region. We encourage researchers to consider and account for mitigating circumstances when selecting optimal deployment sites in the future. The SM recorder was closer to a river and to our campsite than the other recorders, so there was much more background noise. This reduced the detection radius of the recorder as well as the signal-to-noise ratio and therefore also increased the number of false positives.

Deployment site optimization is (rightly) case-specific, making it difficult to generalize protocols but emphasizing the importance of pilot studies.

A fruitful avenue for future research would also be more standardized experiments, such as playbacks, with multiple on-the-ground groups to better ground-truth and compare which specific calls are detected by PAM and in-person observations (van Kuijk et al., 2023). Because roar-shrieks and barks are contagious choruses, and different subgroups engage in these choruses simultaneously (or back and forth to each other), in this study we were not able to directly pinpoint whether a specific chorus we heard in-person is detected at the same time in the PAM recording.

## 5 | CONCLUSION

Our study highlights the potential of PAM as a cost-efficient and scalable tool for biodiversity monitoring, especially in regions where human resources are limited. The use of PAM can provide valuable data on species presence, distribution, and behavior, which can inform conservation efforts and aid in the implementation of effective management strategies. While our study focused specifically on ruffed lemurs in southeastern Madagascar, the workflow presented herein can be adapted and applied to other primate species and regions. We hope that our study will encourage more primatological research using PAM and contribute to the development of more efficient monitoring strategies.

### AUTHOR CONTRIBUTIONS

Carly H. Batist and Andrea L. Baden conceived of and designed the study. Carly H. Batist, Mendrika N. Razafindraibe, and Francois Randriamanantena carried out data collection. Carly H. Batist, Emmanuel Dufourq, and Lorène Jeantet conducted data analysis. Carly H. Batist, Emmanuel Dufourq, and Lorène Jeantet drafted the manuscript, and all authors edited and approved the final version.

### ACKNOWLEDGMENTS

We thank Centre ValBio research station and MICET for their extensive logistical support. C. H. B., M. N. R., and F. R. thank our local guide, Ezafy, and cook/camp lead, Tolotra. Ford, Noro, and the other Mangevo field teams were helpful and accommodating campmates, providing advice and on-the-ground support. C. H. B. thanks the CUNY Graduate Center's Provost Office and Digital Initiatives program for their generous funding. E. D. is supported by a research chairship from the African Institute for Mathematical Sciences, South Africa. This work was carried out with the aid of a grant from the International Development Research Centre, Ottawa, Canada, and with financial support from the Government of Canada, provided through Global Affairs Canada (GAC).

### CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

The full data set of raw recordings is archived on Arbimon (<https://arbimon.rfcx.org/project/mangevo2019/dashboard>) and will be made available upon publication. The training and validation data set along with the model weights are available on Zenodo (<https://zenodo.org/record/7956064>). The training and validation code used for the CNN is available on Github ([https://github.com/AIMS-Research/research\\_z/tree/main/bioacoustics\\_ruffed\\_lemurs](https://github.com/AIMS-Research/research_z/tree/main/bioacoustics_ruffed_lemurs)). The code provided can be used to reproduce the training results (note that due to random weight initialization the results might not be identical). The model weights provided can be used to reproduce the testing results. The software GUI is available on Github (aforementioned link). Installation instructions are provided in a user-manual and demo video. The description of the methodology and the Python code provided on Github will facilitate reproducibility of the methods. A demo of the model being applied to a single audio file is available on Google Co-laboratory ([https://colab.research.google.com/drive/1G\\_zicHNTbJuiXJYsKqRMWktazdX3vx?usp=sharing](https://colab.research.google.com/drive/1G_zicHNTbJuiXJYsKqRMWktazdX3vx?usp=sharing)).

## ORCID

Carly H. Batist  <http://orcid.org/0000-0001-7320-2576>

Emmanuel Dufourq  <http://orcid.org/0000-0002-6986-3408>

Lorène Jeantet  <http://orcid.org/0000-0001-7317-3154>

Andrea L. Baden  <http://orcid.org/0000-0002-4722-0532>

## REFERENCES

- Abadi, M. (2016). TensorFlow: Learning functions at scale, *Proceedings of the 21st ACM SIGPLAN International Conference on Functional Programming* (p. 1).
- Altmann, J. (1974). Observational study of behavior: Sampling methods. *Behaviour*, 49(3–4), 227–266.
- Mac Aodha, O., Gibb, R., Barlow, K. E., Browning, E., Firman, M., Freeman, R., Harder, B., Kinsey, L., Mead, G. R., Newson, S. E., Pandourski, I., Parsons, S., Russ, J., Szodoray-Paradi, A., Szodoray-Paradi, F., Tilova, E., Girolami, M., Brostow, G., & Jones, K. E. (2018). Bat detective-deep learning tools for bat acoustic signal detection. *PLoS Computational Biology*, 14(3), e1005995.
- Appel, C. L., Lesmeister, D. B., Duarte, A., Davis, R. J., Weldy, M. J., & Levi, T. (2023). Using passive acoustic monitoring to estimate northern spotted owl landscape use and pair occupancy. *Ecosphere*, 14(2), e4421.
- Baden, A. L., Oliveras, J., & Gerber, B. D. (2020). Sex-segregated range use by black-and-white ruffed lemurs (*Varecia variegata*) in Ranomafana National Park, Madagascar. *Folia Primatologica*, 92(1), 12–34.
- Baden, A. L., Webster, T. H., & Kamilar, J. M. (2016). Resource seasonality and reproduction predict fission–fusion dynamics in black-and-white ruffed lemurs (*Varecia variegata*). *American Journal of Primatology*, 78(2), 256–279.
- Baden, A. L., Wright, P. C., Louis, E. E., & Bradley, B. J. (2013). Communal nesting, kinship, and maternal success in a social primate. *Behavioral Ecology and Sociobiology*, 67, 1939–1950.
- Batist, C. H., Razafindraibe, M. N., Randriamanantena, F., & Baden, A. L. (2022). Factors affecting call usage in wild black-and-white ruffed lemurs (*Varecia variegata*) at Mangevo, Ranomafana National Park. *Primates*, 63(1), 79–91.
- Batist, C. H., Razafindraibe, M. N., Randriamanantena, F., & Baden, A. L. (2023). Bioacoustic characterization of the black-and-white ruffed lemur (*Varecia variegata*) vocal repertoire. *Primates*, 64, 621–635.
- Beeby, N., & Baden, A. L. (2021). Seasonal variability in the diet and feeding ecology of black-and-white ruffed lemurs (*Varecia variegata*) in Ranomafana National Park, southeastern Madagascar. *American Journal of Physical Anthropology*, 174(4), 763–775.
- Bergler, C., Smeele, S. Q., Tyndel, S. A., Barnhill, A., Ortiz, S. T., Kalan, A. K., Cheng, R. X., Brinklöv, S., Osiecka, A. N., Tougaard, J., Jakobsen, F., Wahlberg, M., Nöth, E., Maier, A., & Klump, B. C. (2022). ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. *Scientific Reports*, 12(1), 21966.
- Bermant, P. C., Bronstein, M. M., Wood, R. J., Gero, S., & Gruber, D. F. (2019). Deep machine learning techniques for the detection and classification of sperm whale bioacoustics. *Scientific Reports*, 9(1), 12588.
- Bittencourt, E., Vasconcellos, A. S., Sousa-Lima, R. S., Young, R. J., & Duarte, M. H. L. (2023). Acoustic monitoring of black-tufted marmosets in a tropical forest disturbed by mining noise. *Animals: An Open Access Journal from MDPI*, 13(3), 352.
- Bittle, M., & Duncan, A. (2013). A review of current marine mammal detection and classification algorithms for use in automated passive acoustic monitoring. *Proceedings of Acoustics*.
- Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M., & Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: Applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3), 758–767.
- Boullhesen, M., Vaira, M., Barquez, R. M., & Akmentins, M. S. (2021). Evaluating the efficacy of visual encounter and automated acoustic survey methods in anuran assemblages of the Yungas Andean forests of Argentina. *Ecological Indicators*, 127, 107750.
- Bravo Sanchez, F. J., Hossain, M. R., English, N. B., & Moore, S. T. (2021). Bioacoustic classification of avian calls from raw sound waveforms with an open-source deep learning architecture. *Scientific Reports*, 11(1), 15733.
- Britt, A. (2000). Diet and feeding behaviour of the black-and-white ruffed lemur (*Varecia variegata variegata*) in the Betampona Reserve, eastern Madagascar. *Folia Primatologica*, 71(3), 133–141.
- Campos-Cerqueira, M., & Aide, T. M. (2016). Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, 7(11), 1340–1348.
- Cannam, C., Landone, C., & Sandler, M. (2010). Sonic visualiser: An open-source application for viewing, analysing, and annotating music audio files. *Proceedings of the 18th ACM Conference on Multimedia* (pp. 1467–1468).
- Castro, I., De Rosa, A., Priyadarshani, N., Bradbury, L., & Marsland, S. (2019). Experimental test of birdcall detection by autonomous recorder units and by human observers using broadcast. *Ecology and Evolution*, 9(5), 2376–2397.
- Chen, X., Zhao, J., Chen, Y., Zhou, W., & Hughes, A. C. (2020). Automatic standardized processing and identification of tropical bat calls using deep learning approaches. *Biological Conservation*, 241, 108269.
- Clink, D. J., & Klinck, H. (2021). Unsupervised acoustic classification of individual gibbon females and the implications for passive acoustic monitoring. *Methods in Ecology and Evolution*, 12(2), 328–341.
- Crunchant, A.-S., Borchers, D., Kühl, H., & Piel, A. (2020). Listening and watching: Do camera traps or acoustic sensors more efficiently detect wild chimpanzees in an open habitat? *Methods in Ecology and Evolution*, 11(4), 542–552.
- Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A., & Tscharrntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: A systematic map and user guide. *Ecological Applications*, 29(6), e01954.

- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database, 2009 *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248–255). IEEE.
- Dew, J. L., & Wright, P. (1998). Frugivory and seed dispersal by four species of primates in Madagascar's Eastern Rain Forest. *Biotropica*, 30(3), 425–437.
- Digby, A., Towsey, M., Bell, B. D., & Teal, P. D. (2013). A practical comparison of manual and autonomous methods for acoustic monitoring. *Methods in Ecology and Evolution*, 4(7), 675–683.
- Dufourq, E., Batist, C., Foquet, R., & Durbach, I. (2022). Passive acoustic monitoring of animal populations with transfer learning. *Ecological Informatics*, 70, 101688.
- Dufourq, E., Durbach, I., Hansford, J. P., Hoepfner, A., Ma, H., Bryant, J. V., Stender, C. S., Li, W., Liu, Z., Chen, Q., Zhou, Z., & Turvey, S. T. (2021). Automated detection of Hainan gibbon calls for passive acoustic monitoring. *Remote Sensing in Ecology and Conservation*, 7(3), 475–487.
- Enari, H., Enari, H. S., Okuda, K., Maruyama, T., & Okuda, K. N. (2019). An evaluation of the efficiency of passive acoustic monitoring in detecting deer and primates in comparison with camera traps. *Ecological Indicators*, 98, 753–762.
- Federman, S., Dornburg, A., Daly, D. C., Downie, A., Perry, G. H., Yoder, A. D., Sargis, E. J., Richard, A. F., Donoghue, M. J., & Baden, A. L. (2016). Implications of lemuriform extinctions for the Malagasy flora. *Proceedings of the National Academy of Sciences*, 113(18), 5041–5046.
- Geissmann, T., & Mutschler, T. (2006). Diurnal distribution of loud calls in sympatric wild indris (*Indri indri*) and ruffed lemurs (*Varecia variegata*): Implications for call functions. *Primates*, 47, 393–396.
- Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods in Ecology and Evolution*, 10(2), 169–185.
- Glander, K. E. (1993). Capture and marking techniques for arboreal primates. In L. G. Campillo, G. P. Montuy, W. M. Contreras Sanchez, & S. L. Arriaga Weiss (Eds.), *Estudios Primatológicos En Mexico* (pp. 299–304). Universidad Juarez Autonoma de Tabasco.
- Grill, T., & Schlüter, J. (2017). Two convolutional neural networks for bird detection in audio signals, 2017 *25th European Signal Processing Conference (EUSIPCO)* (pp. 1764–1768). IEEE.
- Gupta, G., Kshirsagar, M., Zhong, M., Gholami, S., & Ferrer, J. L. (2021). Comparing recurrent convolutional neural networks for large scale bird species classification. *Scientific Reports*, 11(1), 17085.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778).
- Heinicke, S., Kalan, A. K., Wagner, O. J. J., Mundry, R., Lukashevich, H., & Kühl, H. S. (2015). Assessing the performance of a semi-automated acoustic monitoring system for primates. *Methods in Ecology and Evolution*, 6(7), 753–763.
- Hofer, S., McKnight, D. T., Allen-Ankins, S., Nordberg, E. J., & Schwarzkopf, L. (2023). Passive acoustic monitoring in terrestrial vertebrates: A review. *Bioacoustics*, 32, 506–531. <https://doi.org/10.1080/09524622.2023.2209052>.
- Hutschenreiter, A., Kalan, A. K., Bonilla Moheno, M., Morales Mávila, J. E., Mandujano, S., Briseño Jaramillo, M., Spaan, D., & Aureli, F. (2022). Spider monkeys (*Ateles geoffroyi*) habituate to anthropogenic pressure in a low-impact tourism area: Insights from a multi-method approach. *International Journal of Primatology*, 43(5), 946–964.
- Irwin, M. T., Johnson, S. E., & Wright, P. C. (2005). The state of lemur conservation in south-eastern Madagascar: Population and habitat assessments for diurnal and cathemeral lemurs using surveys, satellite imagery and GIS. *Oryx*, 39(2), 204–218.
- Jeantet, L., & Dufourq, E. (2023). Improving deep learning acoustic classifiers with contextual information for wildlife monitoring. *Ecological Informatics*, 77, 102256.
- K. Lisa Yang Center for Conservation Bioacoustics at the Cornell Lab of Ornithology. (2023). *Raven Pro: Interactive sound analysis software (Version 1.6.4) [Computer software]*. The Cornell Lab of Ornithology. <https://ravensoundsoftware.com/>
- Kahl, S., Wood, C. M., Eibl, M., & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61, 101236.
- Kalan, A. K., Mundry, R., Wagner, O. J. J., Heinicke, S., Boesch, C., & Kühl, H. S. (2015). Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. *Ecological Indicators*, 54, 217–226.
- Kalan, A. K., Piel, A. K., Mundry, R., Wittig, R. M., Boesch, C., & Kühl, H. S. (2016). Passive acoustic monitoring reveals group ranging and territory use: A case study of wild chimpanzees (*Pan troglodytes*). *Frontiers in Zoology*, 13(1), 34.
- van Kuijk, S. M., O'Brien, S., Clink, D. J., Blake, J. G., & Di Fiore, A. (2023). Automated detection and detection range of primate duets: A case study of the red titi monkey (*Plecturocebus discolor*) using passive acoustic monitoring. *Frontiers in Ecology and Evolution*, 11, 1173722.
- Kvsn, R. R., Montgomery, J., Garg, S., & Charleston, M. (2020). Bioacoustics data analysis—A taxonomy, survey and open challenges. *IEEE Access*, 8, 57684–57708.
- Lauha, P., Somervuo, P., Lehtikainen, P., Geres, L., Richter, T., Seibold, S., & Ovaskainen, O. (2022). Domain-specific neural networks improve automated bird sound recognition already with small amount of local data. *Methods in Ecology and Evolution*, 13(12), 2799–2810.
- Lawson, J., Rizos, G., Jasinghe, D., Whitworth, A., Schuller, B., & Banks-Leite, C. (2023). Automated acoustic detection of Geoffroy's spider monkey highlights tipping points of human disturbance. *Proceedings of the Royal Society B*, 290(1995), 20222473.
- Leach, E. C., Burwell, C. J., Ashton, L. A., Jones, D. N., & Kitching, R. L. (2016). Comparison of point counts and automated acoustic monitoring: Detecting birds in a rainforest biodiversity survey. *Emu-Austral Ornithology*, 116(3), 305–309.
- Louis, E. E., Sefczek, T. M., Raharivololona, B., King, T., Morelli, T. L., & Baden, A. (2020). *Varecia variegata*. The IUCN Red List of Threatened Species.
- Macedonia, J. M. (1990). What is communicated in the antipredator calls of lemurs: Evidence from playback experiments with ring-tailed and ruffed lemurs. *Ethology*, 86(3), 177–190.
- Macedonia, J. M., & Taylor, L. L. (1985). Subspecific divergence in a loud call of the ruffed lemur (*Varecia variegata*). *American Journal of Primatology*, 9(4), 295–304.
- Markolf, M., Zinowsky, M., Keller, J. K., Borys, J., Cillov, A., & Schülke, O. (2022). Toward passive acoustic monitoring of lemurs: Using an affordable open-source system to monitor phaner vocal activity and density. *International Journal of Primatology*, 43(3), 409–433.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). *Ibrosa: Audio and music signal analysis in python. Proceedings of the 14th python in science conference*, 8, 18–25.
- Mellinger, D., Stafford, K., Moore, S., Dziak, R., & Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans. *Oceanography*, 20(4), 36–45.
- Melo, I., Llusia, D., Bastos, R. P., & Signorelli, L. (2021). Active or passive acoustic monitoring? Assessing methods to track anuran communities in tropical savanna wetlands. *Ecological Indicators*, 132, 108305.
- Morelli, T. L., Smith, A. B., Mancini, A. N., Balko, E. A., Borgerson, C., Dolch, R., Farris, Z., Federman, S., Golden, C. D., Holmes, S. M., Irwin, M., Jacobs, R. L., Johnson, S., King, T., Lehman, S. M., Louis, E. E., Murphy, A., Randriahaingona, H. N. T., Randrianarimanana, H. L. L., ... Baden, A. L. (2020). The fate of



- Madagascar's rainforest habitat. *Nature Climate Change*, 10(1), 89–96.
- Moses, K. L., & Semple, S. (2011). Primary seed dispersal by the black-and-white ruffed lemur (*Varecia variegata*) in the Manombo forest, south-east Madagascar. *Journal of Tropical Ecology*, 27(5), 529–538.
- Do Nascimento, L. A., Pérez-Granados, C., & Beard, K. H. (2021). Passive acoustic monitoring and automatic detection of diel patterns and acoustic structure of howler monkey roars. *Diversity*, 13(11), 566.
- Paumen, Y., Mälzer, M., Alipek, S., Moll, J., Lüdtkke, B., & Schauer-Weissahn, H. (2022). Development and test of a bat calls detection and classification method based on convolutional neural networks. *Bioacoustics*, 31(5), 505–516.
- Pereira, M. E., Seeligson, M. L., & Macedonia, J. M. (1988). The behavioral repertoire of the black-and-white ruffed lemur, *Varecia variegata variegata* (Primates: Lemuridae). *Folia Primatologica*, 51(1), 1–32.
- Pérez-Granados, C., & Schuchmann, K.-L. (2021). Passive acoustic monitoring of the diel and annual vocal behavior of the Black and Gold Howler Monkey. *American Journal of Primatology*, 83(3), e23241.
- Petter, J. J., & Charles-Dominique, P. (1979). Vocal communication in prosimians. In G. Doyle, & R. Martin (Eds.), *The study of prosimian behavior* (pp. 247–305). Academic Press.
- Piel, A. K. (2018). Temporal patterns of chimpanzee loud calls in the Issa Valley, Tanzania: Evidence of nocturnal acoustic behavior in wild chimpanzees. *American Journal of Physical Anthropology*, 166(3), 530–540.
- Piel, A. K., Crunchant, A., Knot, I. E., Chalmers, C., Fergus, P., Mulero-Pázmány, M., & Wich, S. A. (2021). Noninvasive technologies for primate conservation in the 21st century. *International Journal of Primatology*, 43, 133–167.
- Rasmussen, D. T. (1985). A comparative study of breeding seasonality and litter size in eleven taxa of captive lemurs (*Lemur* and *Varecia*). *International Journal of Primatology*, 6, 501–517.
- Ratsimbazafy, J. H. (2002). *On the brink of extinction and the process of recovery: Responses of black-and-white ruffed lemurs (Varecia variegata variegata) to disturbance in Manombo Forest, Madagascar*. State University of New York at Stony Brook.
- Ravaglia, D., Ferrario, V., De Gregorio, C., Carugati, F., Raimondi, T., Cristiano, W., Torti, V., Hardenberg, A. V., Ratsimbazafy, J., Valente, D., Giacomini, C., & Gamba, M. (2023). There you are! automated detection of indris' songs on features extracted from passive acoustic recordings. *Animals: An Open Access Journal from MDPI*, 13(2), 241.
- Roemer, C., Julien, J.-F., Ahoudji, P. P., Chassot, J. M., Genta, M., Colombo, R., Botto, G., Negreira, C. A., Djossa, B. A., Ing, R. K., Hassanin, A., Rufay, V., Uriot, Q., Participants, V. C., & Bas, Y. (2021). An automatic classifier of bat sonotypes around the world. *Methods in Ecology and Evolution*, 12(12), 2432–2444.
- Ruan, W., Wu, K., Chen, Q., & Zhang, C. (2022). ResNet-based bioacoustics presence detection technology of Hainan gibbon calls. *Applied Acoustics*, 198, 108939.
- Schwitzer, C., Mittermeier, R. A., Johnson, S. E., Donati, G., Irwin, M., Peacock, H., Ratsimbazafy, J., Razafindramanana, J., Louis, E. E., Chikhi, L., Colquhoun, I. C., Tinsman, J., Dolch, R., LaFleur, M., Nash, S., Patel, E., Randrianambinina, B., Rasolofoharivelo, T., & Wright, P. C. (2014). Averting lemur extinctions amid Madagascar's political crisis. *Science*, 343(6173), 842–843.
- Spillmann, B., van Noordwijk, M. A., Willems, E. P., Mitra Setia, T., Wipfli, U., & van Schaik, C. P. (2015). Validation of an acoustic location system to monitor Bornean orangutan (*Pongo pygmaeus wurmbii*) long calls. *American Journal of Primatology*, 77(7), 767–776.
- Steinfath, E., Palacios-Muñoz, A., Rottschäfer, J. R., Yuezak, D., & Clemens, J. (2021). Fast and accurate annotation of acoustic signals with deep neural networks. *eLife*, 10, e68837.
- Stowell, D. (2022). Computational bioacoustics with deep learning: A review and roadmap. *PeerJ*, 10, e13152.
- Sugai, L. S. M., Silva, T. S. F., Ribeiro, J. W., & Llusia, D. (2019). Terrestrial passive acoustic monitoring: Review and perspectives. *BioScience*, 69(1), 15–25.
- Thompson, M. E., Schwager, S. J., Payne, K. B., & Turkalo, A. K. (2010). Acoustic estimation of wildlife abundance: Methodology for vocal mammals in forested habitats. *African Journal of Ecology*, 48(3), 654–661.
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., van Langevelde, F., Burghardt, T., Kays, R., Klinck, H., Wikelski, M., Couzin, I. D., van Horn, G., Crofoot, M. C., Stewart, C. V., & Berger-Wolf, T. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communications*, 13(1), 792.
- Turner, M. E., & Harrenstien, L. (1985). Loud calls of the ruffed lemur, *Varecia variegata*. *Journal of the Elisha Mitchell Scientific Society*, 101(1), 1–8.
- Tzirakis, P., Shiarella, A., Ewers, R., & Schuller, B. W. (2020). Computer audition for continuous rainforest occupancy monitoring: The case of Bornean gibbons' call detection. *Interspeech*, 2020, 1211–1215.
- Vasey, N., & Borgerson, C. (2009). Social and ecological factors associated with temporary reproductive collapse in cooperatively breeding red variegated lemurs (*Varecia rubra*) on the Masoala Peninsula, Madagascar. *American Journal of Physical Anthropology*, 261.
- Vieilledent, G., Grinand, C., Rakotomalala, F. A., Ranaivosoa, R., Rakotoarijaona, J.-R., Allnutt, T. F., & Achard, F. (2018). Combining global tree cover loss data with historical national forest cover maps to look at six decades of deforestation and forest fragmentation in Madagascar. *Biological Conservation*, 222, 189–197.
- Vu, T. T., & Tran, L. M. (2019). An application of autonomous recorders for gibbon monitoring. *International Journal of Primatology*, 40(2), 169–186.
- Wang, X., Wen, S., Niu, N., Wang, G., Long, W., Zou, Y., & Huang, M. (2022). Automatic detection for the world's rarest primates based on a tropical rainforest environment. *Global Ecology and Conservation*, 38, e02250.
- Wang, Y., Ye, J., & Borchers, D. L. (2022). Automated call detection for acoustic surveys with structured calls of varying length. *Methods in Ecology and Evolution*, 13(7), 1552–1567.
- Wood, C. M., Barceinas Cruz, A., & Kahl, S. (2023). Pairing a user-friendly machine-learning animal sound detector with passive acoustic surveys for occupancy modeling of an endangered primate. *American Journal of Primatology*, 85, e23507.
- Wright, P. C., Erhart, E. M., Tecot, S., Baden, A. L., Arrigo-Nelson, S. J., Herrera, J., & Zohdy, S. (2012). Long-term lemur research at Centre ValBio, Ranomafana National Park, Madagascar. In P. M. Kappeler, & D. P. Watts (Eds.), *Long-Term Field Studies of Primates* (pp. 67–100). Springer.
- Zaluar, M. T., Tardin, R., Llusia, D., Niemeyer, J., Ribeiro, M. C., & Vale, M. M. (2022). Impact of invasive marmosets (Primates, Callitrichidae) on bird acoustic diversity in a large neotropical urban forest. *Biological Invasions*, 24(6), 1725–1737.
- Zambolli, A. H., Manzano, M. C. R., Honda, L. K., Rezende, G. C., & Culot, L. (2023). Performance of autonomous recorders to detect a cryptic and endangered primate species, the black lion-tamarin (*Leontopithecus chrysopygus*). *American Journal of Primatology*, 85(1), e23454.
- Zhong, M., LeBien, J., Campos-Cerqueira, M., Dodhia, R., Lavista Ferres, J., Velez, J. P., & Aide, T. M. (2020). Multispecies bioacoustic classification using transfer learning of deep convolutional neural networks with pseudo-labeling. *Applied Acoustics*, 166, 107375.
- Zhong, M., Taylor, R., Bates, N., Christey, D., Basnet, H., Flippin, J., Palkovitz, S., Dodhia, R., & Lavista Ferres, J. (2021). Acoustic detection of regionally rare bird species through deep convolutional neural networks. *Ecological Informatics*, 64, 101333.

- Zimmer, W. M. X. (2011). *Passive acoustic monitoring of cetaceans*. Cambridge University Press.
- Zwerts, J. A., Treep, J., Kaandorp, C. S., Meewis, F., Koot, A. C., & Kaya, H. (2021). *Introducing a central African primate vocalisation dataset for automated species classification*. ArXiv Preprint ArXiv:2101.10390.

### SUPPORTING INFORMATION

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

**How to cite this article:** Batist, C. H., Dufourq, E., Jeantet, L., Razafindraibe, M. N., Randriamanantena, F., & Baden, A. L. (2024). An integrated passive acoustic monitoring and deep learning pipeline for black-and-white ruffed lemurs (*Varecia variegata*) in Ranomafana National Park, Madagascar. *American Journal of Primatology*, e23599. <https://doi.org/10.1002/ajp.23599>