Received: 4 August 2025 | Accepted: 5 November 2025

DOI: 10.1111/2041-210x.70213

BRITISH ECOLOGICAL SOCIETY Methods in Ecology and Evolution

RESEARCH ARTICLE

Introducing TropiCam-AI: A taxonomically flexible automated classifier of Neotropical arboreal mammals and birds from camera-trap data

Andrea Zampetti[1,2] | Luca Santini[2] | Iago Ferreiro-Arias[1] | Laura Paltrinieri[1] | Iván Ortiz[1] | Brayan A. Cedeño-Panchez[1] | Christophe Baltzinger[3] | Christopher Beirne[4,5] | Mark T. Bowler[6] | Pierre-Michel Forget[7] | Eric Guilbert[7] | Yvonne J. M. Kemp[4,5] | Carlos A. Peres[8,9] | Andressa B. Scabin[8] | Andrew Whitworth[4,5] | Ana Benítez-López[1]

[1]Department of Biogeography and Global Change, National Museum of Natural Sciences (MNCN-CSIC), Madrid, Spain; [2]Department of Biology and Biotechnologies “Charles Darwin”, Sapienza University of Rome, Rome, Italy; [3]Research Unit Forest Ecosystems, INRAE Val de Loire, Nogent-sur-Vernisson, France; [4]Osa Conservation, Washington, District of Columbia, USA; [5]Osa Conservation Campus, Puntarenas, Costa Rica; [6]School of Engineering, Arts, Science and Technology, University of Suffolk, Ipswich, UK; [7]UMR MECADEV, Muséum National d'Histoire Naturelle-CNRS, Adaptations du Vivant, Paris, France; [8]Instituto Juruá, Manaus, Brazil and [9]School of Environmental Sciences, University of East Anglia, Norwich, UK

Correspondence
Andrea Zampetti
Email: andrea.zampetti@uniroma1.it

Funding information
MCIN/AEI/10.13039/501100011033, “NextGenerationEU”/PRTR, Grant/Award Number: RYC2021-031737-I; MCIN/AEI/10.13039/501100011033/FEDER, EU, Grant/Award Number: PID2022-138272NA-I00

Handling Editor: Giovanni Strona

Abstract

1. Rapid, accurate assessment of arboreal vertebrates in tropical forests remains a bottleneck for large-scale biodiversity monitoring, due to the challenges and effort associated with traditional survey methods. To bridge this gap, arboreal camera-trapping is emerging as a promising way to observe otherwise elusive species, opening new avenues to advance behavioural ecology, community ecology and conservation biology. Yet, unlike ground-based camera-trapping, which has greatly benefited from machine learning innovations for automated species classification, equivalent tools for arboreal wildlife are almost non-existent.
2. Here, we introduce TropiCam-AI, the first algorithm for automated species classification of Neotropical arboreal mammals and birds. We trained a deep learning architecture to recognize 84 taxa (63 species, 13 genera, 5 families and 3 orders), using a diverse set of camera-trap databases from Brazil (77,657 images), Peru (48,670 images), Costa Rica (44,857 images) and French Guiana (18,221 images). We also included in the training phase citizen science images from the iNaturalist platform (53,960 images) to increase generalizability and taxonomic coverage. Finally, we implemented a post-hoc aggregation strategy to further refine uncertain predictions by classifying them at higher taxonomic levels.
3. TropiCam-AI reaches state-of-the-art testing accuracy of 95.0%, with the majority of taxa (50 out of 84) achieving >90% precision and recall, and 36 exceeding 95%.

When allowing our model to return uncertain predictions at higher taxonomic ranks, predictive accuracy increased significantly, with a net gain of +4.3% in accuracy. Thanks to the carefully tuned data augmentation from iNaturalist, our model covers all 24 genera of New World monkeys and reliably classifies a wide array of other arboreal vertebrates.

4. TropiCam-AI is available to use on the no-code AddaxAI platform, enabling a smooth and rapid uptake by a broad range of professionals working in the Neotropics. By reducing reliance on labour-intensive manual review of camera-trap data, TropiCam-AI empowers ecologists to accelerate data processing and ultimately enable effective assessment of Neotropical arboreal wildlife. This versatility can support not only targeted conservation actions but also fundamental ecological research in one of the world's most biodiverse yet under-surveyed ecosystems.

KEYWORDS

arboreal, artificial intelligence, camera-traps, canopy, machine learning, Neotropics, primates, species classification

1 | INTRODUCTION

Camera-traps have emerged over the last few decades as a cornerstone of wildlife monitoring, offering a non-invasive, cost-effective and continuous means of surveying vertebrate communities across broad spatial and temporal scales (Silveira et al., 2003). Mainly employed for ground-based studies, they are particularly effective for detecting nocturnal or elusive species, which would be increasingly difficult to study through direct observation (Moore et al., 2021). Because camera-traps are able to support multi-species survey programmes, they have become a staple technique to monitor terrestrial vertebrates in the context of behavioural studies (Caravaggi et al., 2017), as well as for the estimation of occupancy (MacKenzie et al., 2017), and the evaluation of species richness (Ahumada et al., 2011; Tobler et al., 2015). Due to these valuable properties, in recent times researchers have also begun adapting these systems to the forest canopy (Coutant et al., 2022; Moore et al., 2021; Séguigne et al., 2022; Zhu et al., 2022), deploying arboreal camera-traps to access the vertical stratum of tropical rainforests and capture data on entirely different suites of species (Bowler et al., 2017; Whitworth et al., 2016).

However, a persistent challenge of camera-trapping studies is the vast amount of data produced, and the subsequent need to manually annotate the thousands (or even millions) of images gathered in the field. This introduces a clear gap between data acquisition and efficient extraction of meaningful ecological metrics. Such a delay defies the purpose of readily tracking changes in faunal communities' composition and persistence in a context of emerging threats such as habitat loss, overhunting, and rapidly changing climate, that can elicit significant impacts on natural environments (Tilman et al., 2017). One possible solution to this barrier is the integration of artificial intelligence (AI) to automate the data sorting and classification process. In recent years, technical advances in the field of computer vision have enabled the conceptualization of machine learning algorithms loosely inspired by the visual cortex in the mammalian brain (Hu et al., 2015), that are able to perform image-based tasks after being instructed on a labelled dataset. Convolutional neural networks (hereafter, CNNs) are a specific class of deep learning algorithms that have gained widespread visibility due to their high accuracy and flexibility in extracting information from visual data such as images (He et al., 2016; Simonyan, 2014). Due to these properties, they have been successfully applied to the analysis of camera-trap data to carry out tasks such as animal detection, counting and taxonomic classification (Beery et al., 2019; Norouzzadeh et al., 2021; Tabak et al., 2020), in some cases even outperforming expert-based labelling and reducing manual effort by up to two orders of magnitude (Fennell et al., 2022; Willi et al., 2019). Consequently, ecologists are now starting to explore the feasibility of applying machine learning classifiers directly to camera-trap analyses to obtain ecological metrics of interest such as activity patterns (Mitterwallner et al., 2024; Whytock et al., 2021), species richness and occupancy (Whytock et al., 2021), population density (Zampetti et al., 2024) and species interactions (Villalva & Jordano, 2025) without (or with minimal) human supervision. At present, a variety of CNNs for species classification have been developed: from local algorithms trained on a geographically restricted subset of species (Rigoudy et al., 2023; Tabak et al., 2020), to large web-based platforms aimed at reaching global taxonomic coverage (Ahumada et al., 2020).

Despite these advances, arboreal camera trapping remains severely underrepresented compared to AI trained on terrestrial images. This is particularly acute in the Neotropical realm, where the complex three-dimensional forest structure and exceptionally rich arboreal vertebrate fauna (Honda et al., 2025; Myers et al., 2000) underscore the difficulty of monitoring such biodiverse assemblages, which are of critical ecological and conservation relevance (Jenkins et al., 2013; Pillay et al., 2022). Arboreal vertebrate communities play a key role in fundamental ecological processes such as seed dispersal, with up to 90% of tropical rainforest plant species being consumed by primates, small mammals, birds and bats (Howe & Smallwood, 1982). Notably, large vertebrates such as primates serve as long-distance dispersers, helping to maintain structural integrity and biodiversity in plant communities (Link & Di Fiore, 2006; Nathan & Muller-Landau, 2000). Their ability to handle large-sized fruits, paired with their greater energetic requirements, enables them to mobilize and process a larger number of seeds compared to other small and medium-sized species, making them key agents in the seed dispersal of a wide array of plant species (Link & Di Fiore, 2006; Peres & van Roosmalen, 2002). Yet these communities are especially vulnerable to hunting and habitat loss (Whitworth et al., 2019). Given the urgent need to detect and respond to acute declines driven by environmental and anthropogenic stressors, the development of targeted AI-driven monitoring and analytical tools is indispensable for rapidly assessing species distributions, richness patterns, and the conservation status of arboreal species assemblages.

Here, we present TropiCam-AI, a flexible and effective automated classifier of Neotropical arboreal mammal and bird species. Trained on an extensive camera-trapping dataset and augmented with citizen science images, our algorithm is able to identify 84 different taxa of arboreal mammals and birds (62 species, 14 genera, 5 families and 3 orders), and notably, covers all platyrrhine primate genera found in the Neotropics. TropiCam-AI provides a free and user-friendly tool that can directly assist ecologists and conservation practitioners in rapidly annotating camera-trap images and videos in a semi-automated way. It is available on the AddaxAI platform (https://addaxdatascience.com/addaxai/; Van Lunteren, 2023), an interactive interface designed to make it accessible to researchers irrespective of technical skills, hence ensuring a smooth and swift uptake by a broad range of professionals working in the Neotropics. Furthermore, to address the 'black box' nature of CNNs while supporting flexibility and interpretability of predictions, we integrated a recently described approach to predict on multiple taxonomic levels without the need for (computationally expensive) hierarchical neural networks. This approach increases model performance on challenging datasets and can be easily integrated into pre-existing CNNs from other ecological domains.

2 | MATERIALS AND METHODS

2.1 | Training datasets

As the accuracy of CNNs is highly dependent on the abundance of the available training data, here we sought to maximize sample size as much as possible, particularly for rare and/or threatened species for which high classification performance is critical in order not to lose valuable information. Thus, we adopted a hybrid approach where the bulk of our dataset was derived from different camera-trap projects in the Neotropics, and then we augmented it using citizen science images obtained from the iNaturalist platform. We mobilized our network of collaborators in the Neotropics to identify all potential partners that have past or ongoing arboreal camera-trapping projects. We collected a total of more than 85 thousand images and 26 thousand videos containing animals across nine sampling sites in Brazil, Peru, Costa Rica and French Guiana. From these, we only selected species that were captured by camera-traps in the canopy. All videos were converted to images by extracting three frames per second (see Table S1 for the final dataset). After manual inspections of the dataset to assess possible misclassifications, we standardized it to a common taxonomy following the latest available version of the Catalogue of Life (Bánki et al., 2025). The final dataset included 63 species (33 primates, 5 marsupials, 3 sloths, 2 anteaters, 4 carnivores, 3 rodents and 13 birds) and 15 higher taxonomy categories (genus, family and order; refer to the Supporting Information for the full list of taxa). The latter were dictated by either too low a sample size at the species level to be effectively learned by our algorithm (e.g. some species had as few as 5 images), or because those images could not be identified at a higher taxonomic resolution by any of the experts that provided the data. We decided to still include them as higher categories because of how commonly they occur in arboreal camera-trapping surveys (e.g. genus *Ramphastos*), or because of their relevance in terms of conservation status (e.g. genus *Cebuella*).

2.1.1 | iNaturalist

Here, we wanted to harness the potential of iNaturalist (iNaturalist, 2025), one of the biggest citizen science platforms for biodiversity data sharing, to augment our camera-trapping dataset in an effort to expand the generalizability and performance of our classifier. We retrieved observations with images for all Platyrrhines and a selection of other arboreal mammals and birds for which iNaturalist images were available either at the species or genus level through the iNaturalist API on May 2024, filtering for the ones that were labelled as 'research grade' (representing the highest quality of observations, for which taxon identity could be confirmed) and that were posted on a Creative Commons (CC) licence. We also leveraged the iNaturalist Open Data, a curated pre-labelled database originally created for Kaggle competitions (Van Horn et al., 2018). We repeated the procedure while accounting for duplicates between the two datasets (iNaturalist API and iNaturalist Open Data). At the end of the process, we were able to acquire 97,942 new images (Table S1), which allowed us to include one additional species (*Callimico goeldii*) and six Platyrrhini genera that were not represented in our camera-trap dataset (*Brachyteles*, *Callithrix*, *Chiropotes*, *Leontopithecus*, *Mico* and *Oedipomidas*). Those were included in the training dataset as categories at the genus level, to meet our set requirements in terms of sample size.

2.2 | Data preprocessing

Previous work has demonstrated the superior efficiency of a two-step classifier in automatically classifying animals in images (Gadot et al., 2024). Here, we adopted the same pipeline. We used MegaDetector v.5 (Beery et al., 2019), a state-of-the-art object detection algorithm already widespread among camera-trap practitioners, to crop animals and get rid of the background information. We retained only cropped images that had a confidence score above 0.8 to minimize eventual false inclusions (Rigoudy et al., 2023). Then, we split each category in the dataset into training, validation and testing (80%–10%–10%): as camera traps often record sequences of images for the same individual, we split at the sequence level instead of at the image level to avoid pseudoreplication. While there is no general consensus on a fixed time-to-independence value (which can be highly context-dependent; Peral et al., 2022), we used a threshold of 15 min, that we deemed appropriate after manual inspection of our data. For a number of sites, we did not have access to date and time metadata, so we adopted a text similarity approach where we clustered similar images by file names using the DBSCAN algorithm (Khan et al., 2014), treating them as sequences and then manually validating them. At the end of the preprocessing phase, our dataset comprised a total of 243,336 images, divided into 191,009 (training), 25,638 (validation) and 26,689 (testing).

2.3 | Machine learning model

2.3.1 | Training and fine-tuning

For our model, we selected the recent ConvNeXt-Base CNN architecture (Liu et al., 2022) to maximize accuracy while maintaining manageable training and inference times. We performed data augmentation on the training set by randomly transforming images through rotation, shift, shear, zoom, brightness adjustment, channel shift range, horizontal and vertical flip. While the latter is usually not common, we deemed it appropriate in an arboreal setting where the third dimension can cause animals to be captured in different vertical orientations. Our training dataset was highly heterogeneous in terms of sample size, with the most abundant category (the Central American spider monkey *Ateles geoffroyi*, $n = 23{,}590$) being more than three orders of magnitude larger than the least abundant (*Callimico goeldii*, $n = 6$). We decided to adopt a custom weighted loss during training, to assign proportionally greater penalties to misclassified minority classes and lower penalties to majority classes. This prevents the model from being biased towards the dominant class, improving performance on rare categories (Lin et al., 2018). We tested four scenarios: no weighting, inverse ($w_i = \frac{1}{f_i}$), square root ($w_i = \frac{1}{\sqrt{f_i}}$), and logarithmic weighting ($w_i = \frac{1}{\log(f_i + 1)}$), where w_i represents the weight assigned to category i and f_i the frequency of category i in the dataset. Square root and logarithmic weighting were selected as they are less extreme than inverse weighting, which could potentially lead to instability and poor generalization (Cui et al., 2019).

As CNNs require careful tuning of multiple hyperparameters that can significantly affect model performance, we chose a Bayesian search optimization approach to identify the most optimal set of hyperparameters that would maximize accuracy (Snoek et al., 2012). We tuned learning rate, optimizer, dropout rate, number of units in the fully connected layer and weighting strategy for the custom loss function (see Table S2 for each parameter's interpretation and range of possible values). After the Bayesian search, we trained the full model with the selected hyperparameters using transfer learning from ConvNeXt-Base, on a high-performance computing cluster with two GPUs (CESGA Supercomputing facilities). A summary of the selected hyperparameters used for final training, and further information on the training and validation process can be found in the Supporting Information S1. All analyses were carried out in Python 3.10.15 using TensorFlow 2.10.1 in a Conda environment.

2.3.2 | Testing and performance metrics

We evaluated our model's performance by calculating standard metrics (accuracy, precision, recall and F1-score), but from here on we will strongly emphasize recall (i.e. the ability to correctly identify all instances of a given species), as we consider false-negatives a greater concern than false-positives in a context where camera-trap data are usually manually validated to extract additional and can be corrected for erroneous inclusions. Although the testing split is not used by the model during training and can be effectively considered out-of-sample, it still comprises fractions of the different data pools on which the model was trained. So, we also carried out an additional assessment using a separate labelled dataset, which was completely held out from the training split. This case study data included 1880 images of arboreal mammals and birds as part of a survey that was conducted from July 2017 to May 2019 in the same area of one of our training datasets (the Médio Juruá region, in the western Brazilian Amazon; Scabin & Peres, 2021), although following a different survey protocol. As often camera-traps record multiple snapshots of the same animal passing in the field of view, it makes more ecological sense to consider the capture event as a whole instead of the single image/video. Following Dussert et al. (2025), we aggregated single predictions by averaging their logits over the corresponding sequence and selecting the category with the highest resulting score. Finally, we also tested whether data augmentation from iNaturalist was effective in increasing model performance: to this end, we considered only the taxa for which we had both camera-trapping and iNaturalist images, and we trained a model with and without iNaturalist data augmentation in the training set. Then we tested the performance of both models on a testing set that comprised only camera-trap images and we calculated the difference between taxon-specific F1-scores (which balances the robustness to false-positive and false-negative errors) between the two trained models, to identify whether and for which taxa, the iNaturalist augmentation strategy enhanced model performance.

2.4 | Taxonomic aggregation strategy

Convolutional neural networks are innate 'black box' models, as they treat the categories on which they are trained as independent and uncorrelated entities. This is quite limiting when analysing biological data, which is highly hierarchical and shows clear patterns of correlation between entities (here, the phylogenetic non-independence between closely related species). Here, we adopted a very simple strategy, recently proposed by Gadot et al. (2024), to leverage phenotypic relatedness and thereby turn potential misclassifications into useful information. This pipeline assumes that closely related species tend to share more phenotypic similarity in morphological traits. While we acknowledge that this assumption does not always hold true in mammals and birds, due for example to the evolutionary convergence leading to unrelated species having similar biological traits (Pigot et al., 2020), in this study we only focused on testing whether this could improve classification performance and generalization when applied to computer vision in wildlife monitoring. We chose a thresholding approach for the output confidence probability, where predictions under a selected threshold would be flagged as 'uncertain'. For those, we aggregated the confidence scores at the higher taxonomic level: for example, every species would be mapped to the corresponding genus, and probabilities for the same genus would be summed. We iteratively repeated the procedure through the selected taxonomy levels (species, genus, family, order, class) until the set threshold was met, and we considered that as the final prediction; if the confidence score was still under the threshold after aggregation at the class level, we flagged the prediction as 'unknown animal'. We also wanted to maximize both flexibility (correct predictions at higher taxonomic levels that would have been otherwise missed at the species level) and taxonomic resolution (the proportion of correct predictions at the species level), as ecological analyses typically require information to be collected at the species level where possible. To this end, we adopted a thresholding approach where we assessed model recall on the testing dataset across a range of threshold values (0.01–0.99), and for each, we quantified the proportion of correct classifications that were predicted at the species level to identify the optimal balance point between the two objectives.

3 | RESULTS

3.1 | Model performance on the testing set

Training accuracy for our best model reached 94.8%. When not adopting the taxonomic aggregation strategy, accuracy on the testing set was 90.7%. However, when allowing the algorithm to predict at higher taxonomic levels with our optimal confidence threshold of 0.75 (Figure S3), testing accuracy increased to 95.0%. Predictive performance increased or remained consistent for almost all taxa (81 out of 84; Figure 1), with the sole exception of two bird categories (*Penelope jacquacu* and *Ramphastos* sp.) for which accuracy marginally decreased (<1%). Focusing on the recall metric, the average overall difference in predictive performance between the model with the taxonomic aggregation strategy and the baseline model across all taxa was +10.8 percentage points (+15.3% for primates, +10.0% for other arboreal mammals and +3.1% for birds). Fifty out of 84 taxa had a recall value >90%, and only two exhibited very low values below 50% (*Pipile cumanensis* and Passeriformes; Figure 1). Recall for 36 taxa exceeded 95%, irrespective of the number of training images: some had a high number of training images (e.g. *Ateles chamek* or *Sapajus apella*, $n>10{,}000$ images), and others a few hundred (e.g. *Cebuella* sp. or *Tamarinus imperator*, $n<200$ images). Our model also showed good discrimination abilities for species belonging to the same genus (e.g. *A. chamek* versus *A. geoffroyi*, testing recalls >95%; *Pithecia hirsuta* versus *P. irrorata*, testing recalls >85%). Some other congeneric species tended to be misclassified to a slightly higher extent at the species level (e.g. *Alouatta sara* misclassified as *A. macconnelli*, testing recall = 79%; *C. albifrons* and *C. olivaceus* misclassified as *C. imitator*, testing recalls = 75%), and were successfully predicted with high confidence at the genus level or higher when applying the taxonomic aggregation strategy. When allowing uncertain predictions to be aggregated, the majority of classifications were still returned at the species level for all categories where it was available, with the exceptions of *Callimico goeldii*, *Cheracebus regulus* and *C. torquatus* that were mainly classified at the order level. When the training dataset was augmented using iNaturalist images, global accuracy marginally increased by 0.42 percentage points, but taxon-specific F1-scores varied greatly. While some taxa (21 out of 50) benefitted from the augmented set (mean and maximum F1-score increase = 5.4% and 22.6%, respectively), others (16 out of 50) did not observe any change and the remaining (13 out of 50) actually decreased in predictive performance (mean and maximum F1-score decrease = −6.2% and −22.2%, respectively). Thus, for our final model, we used data augmentation from the iNaturalist dataset only for those taxa whose ΔF1-score >0 (Figure 2).

3.2 | Model performance on the out-of-sample set

When predicting on the out-of-sample set using a confidence of 0.20 for MegaDetector and 0.75 for the taxonomic aggregation, recall at the sequence level was 95.8% for the object detector and 83.7% for the taxonomic classifier. More than half of the species in the dataset (17 out of 32) had recall values >95%; in fact, for 16 of those, our algorithm perfectly classified all sequences (Figure 3, Table S4). The poorest classification performance was observed for *Cacajao calvus* and *Lagothrix lagotricha*, both of which had an extremely low number of images (respectively, $n=8$ and $n=12$) in which they were only partly visible, and were thus respectively mistaken as *C. albifrons* and *A. chamek*. Also, *Tamarinus mystax* was often misclassified due to the high visual similarity with species from the *Leontocebus* genus. When testing the sensitivity to a higher confidence threshold of 0.95, which may reflect a researcher's desire to ensure better image

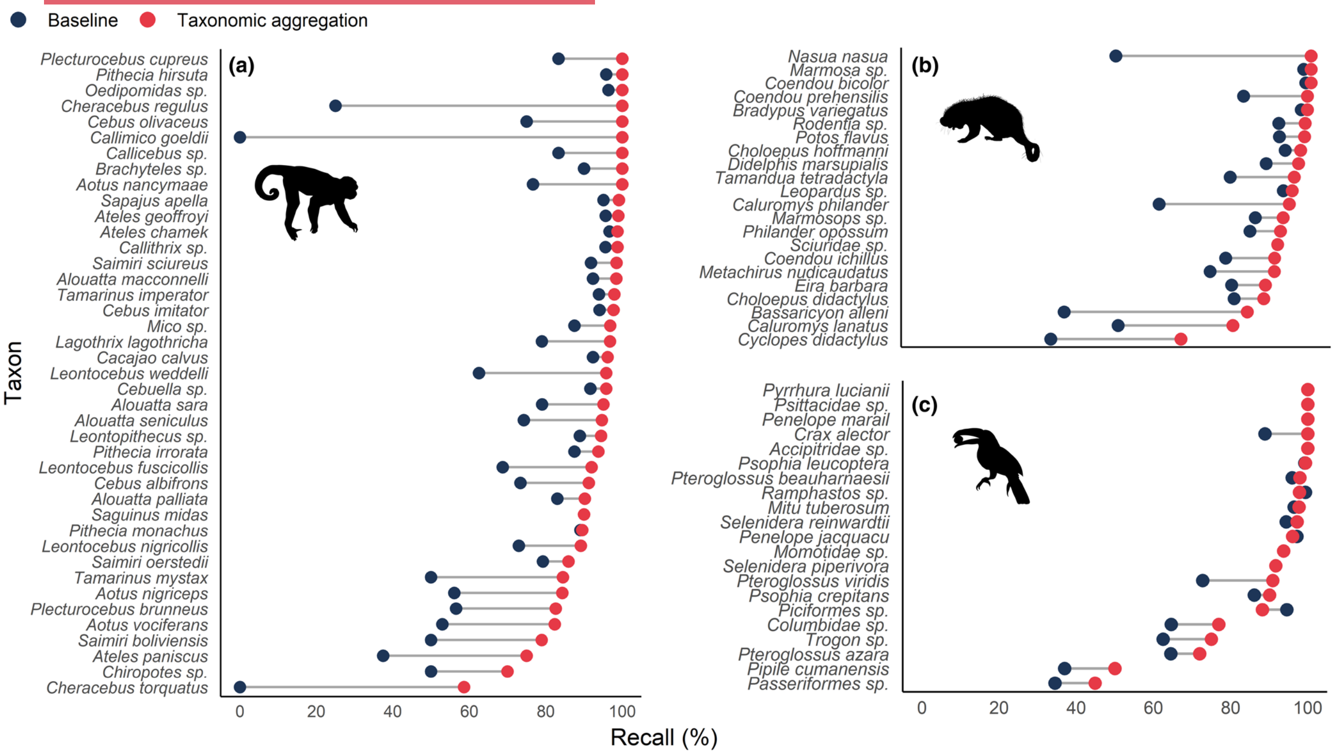

FIGURE 1 Recall parameters for all taxa included in the model, computed on the testing dataset and separated for (a) primates, (b) other arboreal mammals and (c) birds. For each taxon, two recalls are calculated: A baseline where the algorithm is forced to return predictions on the lowest taxonomic level possible (in blue), and one where it is allowed to aggregate uncertain predictions at higher taxonomic levels (in red). Here, the confidence threshold for aggregation was set to 0.75. Silhouettes were downloaded from PhyloPic (https://www.phylopic.org/). The same results in tabular form are presented in Table S3.

inclusion at the expense of taxonomic resolution (Figure S3), we observed a clear increase in performance, both overall (recall = 93.3%) and species-specific (Figure 3, Table S4).

4 | DISCUSSION

TropiCam-AI represents the first deep learning algorithm specifically developed for automated classification of arboreal species in the Neotropics. Our algorithm can identify 84 different taxa of arboreal mammals and birds (62 species, 14 genera, 5 families and 3 orders), with an emphasis on platyrrhine primates, for which taxonomic coverage spans all 24 genera. In addition, thanks to the taxonomic aggregation strategy, our model can adapt to challenging classification conditions by calibrating the output to the confidence of predictions and consequently adjusting the taxonomic resolution, thus providing researchers with a useful tool that goes beyond mere classification and can accommodate the highly hierarchical structure of animal taxonomy. Our model is available and ready to use on the AddaxAI platform (https://addaxdatascience.com/addaxai/), an open-access software specifically designed for seamless and user-friendly usage without any need for programming knowledge. Through that, practitioners can run TropiCam-AI locally on their machines on both images and videos, optionally annotate and manually validate their results, and export them in a number of formats, including an option for compatible integration in the popular Timelapse software (https://timelapse.ucalgary.ca/) for camera-trap data annotation.

To the best of our knowledge, TropiCam-AI is the first species recognition model specifically trained for Neotropical mammals and birds in an arboreal setting. Lately, there has been some effort to bring AI advancements to analyse camera-trapping datasets in Neotropical forests: for example, the Pytorch-Wildlife initiative offers an algorithm trained on 36 genera of mammals and birds from the Amazon rainforest (Hernandez et al., 2024), but it is limited to ground-dwelling animals. The AddaxAI platform (Van Lunteren, 2023) also hosts two models developed by the San Diego Zoo Wildlife Alliance that can recognize 42 and 53 ground-dwelling mammals and birds from the Peruvian Amazon and Peruvian Andes, respectively. The recently released SpeciesNet algorithm (Gadot et al., 2024), which is the AI model behind the Wildlife Insights platform (Ahumada et al., 2020), can recognize some arboreal species from the Neotropics, but is limited both in sampling size and taxonomic coverage for this ecological context (e.g. only three primate species have more than 1000 training images). This makes our AI the most comprehensive solution up to date for geographical and taxonomic coverage across the Neotropical region, and the first one created specifically to target monitoring of arboreal mammals and birds. Given the steady increase in recent years of arboreal camera-trapping studies (Moore et al., 2021), we foresee that our work could

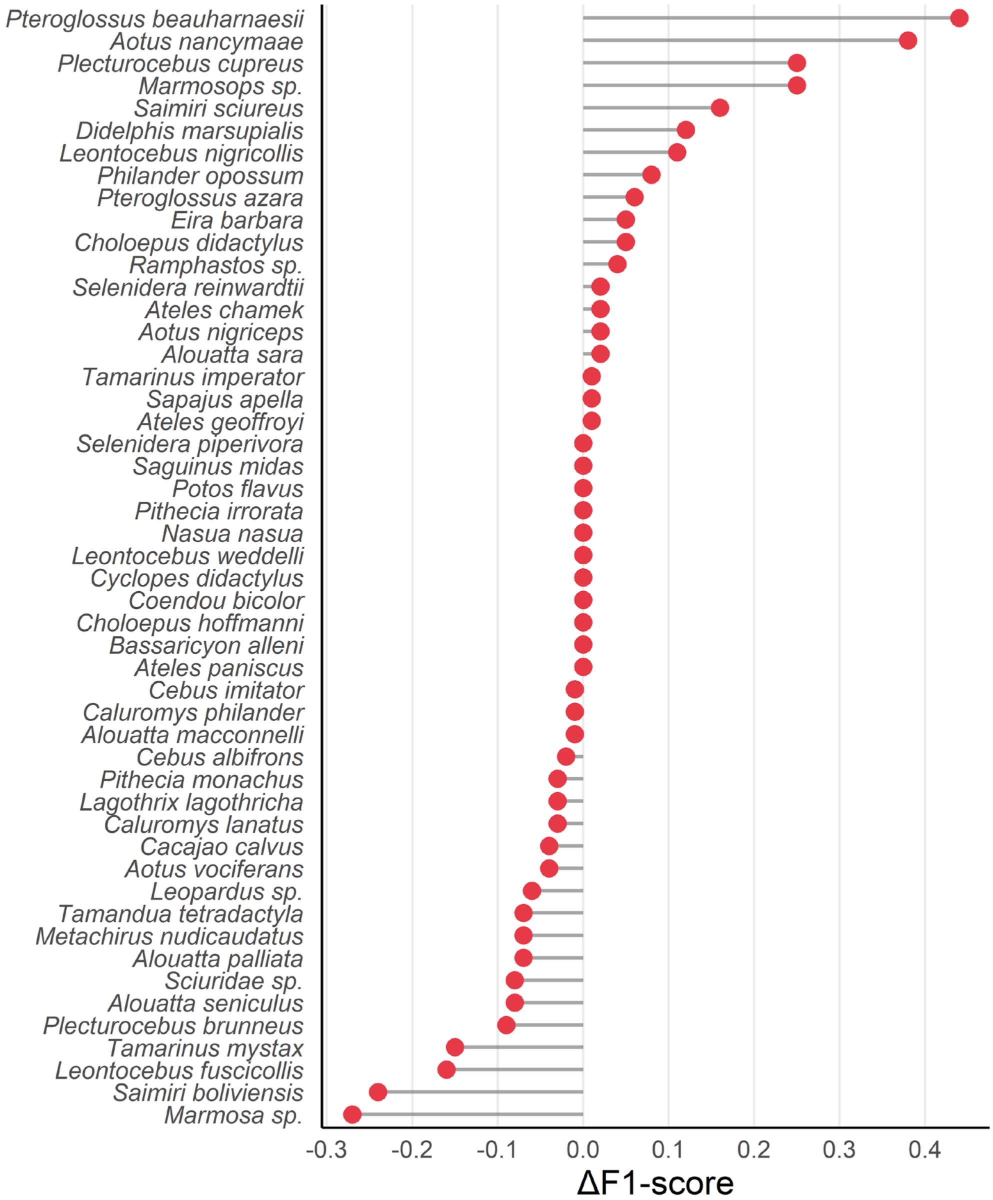

FIGURE 2 Differences in taxon-specific F1-scores calculated when training the model on both camera-trapping and iNaturalist images, computed on a testing set that included only camera-trapping images. For each taxon, values above 0 indicate that the data augmentation strategy led to an increased predicting performance, while balancing both false-negatives (i.e. images of that taxon that were missed by the classifier) and false-positives (i.e. images of another taxon that were mistakenly classified as the current taxon). In the final model, only taxa whose ΔF1-score >0 were augmented with iNaturalist images.

be soon joined by additional efforts that will provide AI solutions for other biogeographic regions where arboreal camera-trapping may be adopted.

Our model showed very promising classification performance across the multiple taxa included, with overall testing accuracies reaching up to 95.0%. We note that our performance is on the higher end for deep learning standards: for example, the recent AI algorithms for ground-dwelling wildlife report out-of-sample accuracies of 94.9% (Norouzzadeh et al., 2018), 93.8% (Tabak et al., 2020), 78.0% (Whytock et al., 2021) and 93.6% (Rigoudy et al., 2023). However, we stress that our model is the first one developed specifically for an arboreal setting, so direct comparison may not be ideal, since the two strata may exhibit different challenges in image classification. Notably, arboreal camera-traps can produce a disproportionately high number of false triggers due to higher exposure to wind and moving foliage (Gregory et al., 2014), which holds particularly true in a rainforest environment. This, paired with the higher complexity and variability of the background, may result in a higher risk of misclassification due to partial occlusion of animals (Tabak et al., 2020). It is critical to take these factors into account when evaluating model performance and, more importantly, practical consequences in a monitoring scenario. Further research is needed to establish proper benchmark evaluations of automated classifiers in an arboreal setting.

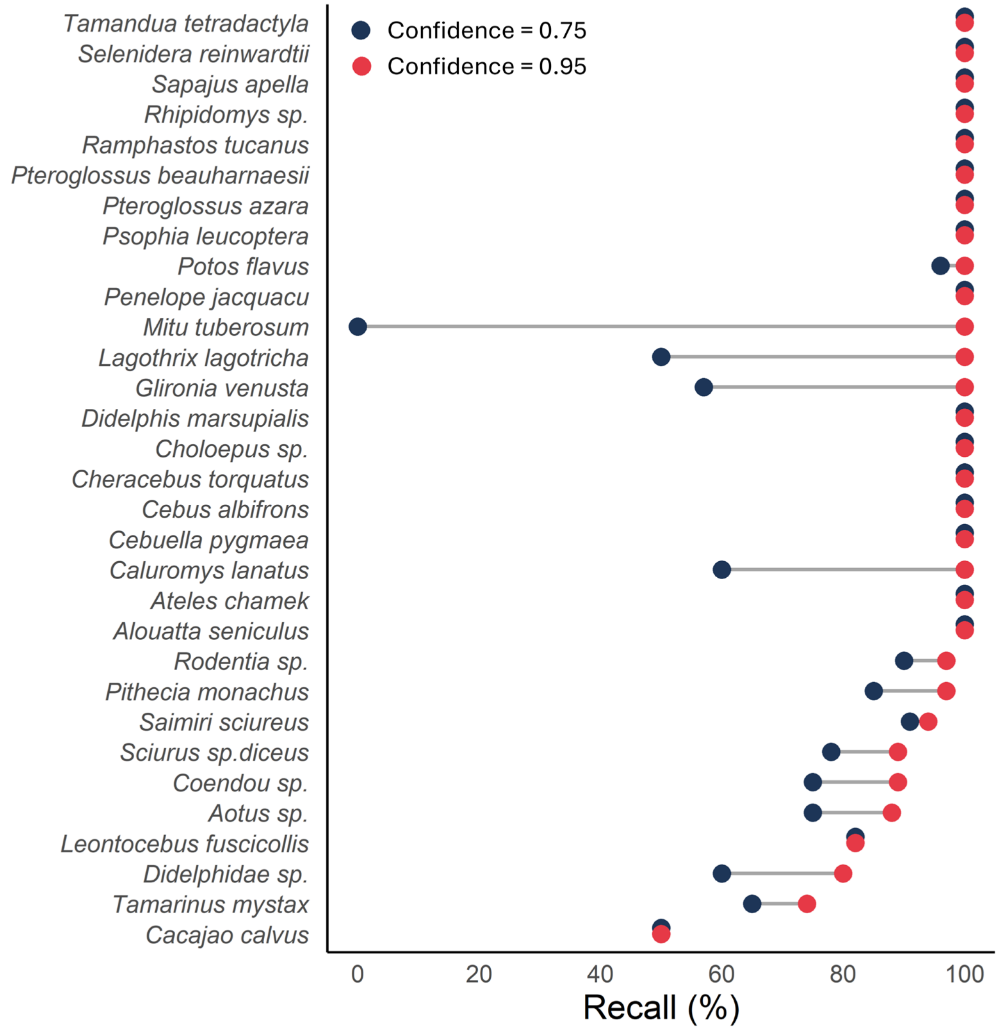

FIGURE 3 Performance metrics of the model's predictions on the out-of-sample Brazilian dataset. Recall is computed at the sequence level, and is reported for each taxon at two confidence thresholds for taxonomic aggregation (0.75 in blue, 0.95 in red).

Model performance at the fine-grained taxonomic level was generally high, with a fairly limited number of taxa ($n=15$) whose recall was <80% (Figure 1). While there is a general consensus in the direct relationship between the number of training images and classification performance (Tabak et al., 2019), our model performed really well even on species with low sample sizes, showing the effectiveness of our weighting strategy to deal with class imbalance. We note that there are a number of ways to tackle this particular issue (e.g. see Rigoudy et al., 2023), but, as decisions in model training and fine-tuning are highly context-dependent (Tuba et al., 2021), we encourage future practitioners to tailor them to their specific needs. Including a suite of strategies directly in the fine-tuning process, as we did here, may represent a valuable alternative against a 'one-strategy-fits-all' approach. Importantly, the biggest contribution to high classification performance across all taxa was reached thanks to the taxonomic aggregation strategy: while recall increased for almost all taxa, in extreme cases the increase was substantial, with for example *Cheracebus torquatus* going from 0.0% to 58.6% recall. In a practical scenario, this implies that those images in which the animal could be otherwise classified with low confidence, would instead be returned at a higher taxonomic level with far greater accuracy (Figure 4). We observed that this strategy pairs really well in a CNN for animal classification, as we noticed misclassifications to be clustered on similar and phylogenetically related species (Supporting Information S2). Given that this post-processing step can be implemented without any required modification to the model itself, it is fairly straightforward to adapt in current algorithms. The Wildlife Insights platform already has it built into their default inference pipeline (Gadot et al., 2024), and the AddaxAI platform will be integrating it in the near future for all its existing models (personal communication from the platform developer and owner), thus further increasing their reported performance and applicability. For TropiCam-AI, we suggest practitioners to experiment with the aggregation threshold based on their specific needs: the default of 0.75 was selected to maximize both recall and the proportion of predictions returned at the species level, but priority can be allocated to either high absolute recall (i.e. fewer images are lost in the classification process, at the cost of lower taxonomic resolution on predictions) or high

FIGURE 4 Example of a challenging image to classify, where a black spider monkey *Ateles chamek* individual is correctly detected but model confidence at species and genus level falls below the aggregation threshold. In this case, the model resorts to the family level (Atelidae, 83.1%) instead of returning an under confident prediction (*Ateles chamek*, 64.3%). The side panel also shows what the model 'sees' when deciding which animal is in this image, with part of the prehensile tail being the key feature that led to the final prediction. Image created using the Grad-CAM technique (Selvaraju et al., 2020).

taxonomic resolution (i.e. more images are lost in the classification process, but a higher proportion of predictions are returned at the species level) by increasing or decreasing the threshold, respectively. Or, if deemed the case, users can decide not to use the taxonomic aggregation at all, and get predictions at the species level as default. We also remind practitioners that when using TropiCam-AI on their data, they are not bound to use the entire list of taxa for predictions: instead, they can deselect species or higher order taxa that are not found in their study area, which will then be excluded from the prediction options (this feature is already built-in and easy to apply in the AddaxAI platform). This way, the probability of confounding, for example, a species with a congeneric that is found in a different geographic region is further diminished.

In this work, we also wanted to test whether the large body of data from large-scale citizen science initiatives can be used to boost domain-specific AI applications such as in a camera-trapping context, a question that has been previously posed (Beery et al., 2020, 2021) although not thoroughly evaluated (but see Shepley et al., 2021). We found that the answer to that question is not straightforward: in our case study, the included taxa reacted differently to data augmentation from iNaturalist. If we had blindly included citizen science images for all the taxa, we would have significantly degraded the performance of our algorithm. This could happen for a variety of reasons: for example, iNaturalist images differ from camera-trap photos in resolution, lighting and framing and inflating the training dataset could cause a shift in input data away from the desired target (a phenomenon known as 'domain shift'; Schneider et al., 2020), which could impact some taxa more than others. Instead, we chose to adopt this strategy only for a subset of our taxa, for which the new images actually improved predictive performance without introducing excessive noise for the neighbouring taxa. We call for caution on this matter, particularly due to the ease with which large citizen science image datasets are now accessible for research purposes (Van Horn et al., 2018; Vendrow et al., 2024), and that are, for example, being leveraged in recent efforts to develop large-scale, cross-domain AI algorithms with massive taxonomic coverage (Stevens et al., 2024). We encourage researchers to keep exploring the possibilities of data integration from such sources, while being mindful of the actual ecological context for which an automated tool is being developed. Our case is particularly complex in terms of the number of closely related (and easy to misclassify) species on which we trained our algorithm: we would be curious to observe if and how data augmentation from iNaturalist could benefit AI performance in a different ecological context, with a more limited subset of species that are not as challenging to distinguish as ours. Perhaps, future research efforts could experiment with existing AI tools trained in a less hyperdiverse setting and see if the strategy we adopted here would yield increased returns.

In summary, TropiCam-AI represents a significant advance in the automated analysis of arboreal camera-trap imagery from the Neotropics. By combining high accuracy with an adjustable taxonomic aggregation strategy, our model not only alleviates the burden of manual labelling but also gives researchers explicit control over the trade-off between model accuracy and adaptability to challenging scenarios. Through its integration into the user-friendly AddaxAI platform, field ecologists and conservation practitioners can leverage deep learning without needing to write a single line of code, thus bridging an important accessibility barrier towards seamless adoption of AI tools in ecological research (Brook et al., 2025). As recent advancements in the literature suggest, more and more research is explicitly showing how integrating AI for species classification in downstream analyses can foster efficient and timely estimation of key ecological metrics, such as diel activity patterns (Mitterwallner et al., 2024; Whytock et al., 2021), species richness and occupancy (Whytock et al., 2021), population density (Zampetti et al., 2024) and community-level interaction networks (Villalva & Jordano, 2025). This holistic approach holds promise in minimizing the need for manual review and providing near real-time generation of high-resolution ecological insights, thereby accelerating both basic research and conservation decision-making. Ultimately, we envision TropiCam-AI as one component in a collaborative, hybrid workflow, in which AI accelerates the sorting and labelling phases of vast image datasets, while expert ecologists retain oversight to ensure ecological interpretations remain grounded in field knowledge. This synergy between machine intelligence and human expertise promises to transform large-scale arboreal biodiversity monitoring, enabling more rapid, reproducible and cost-effective assessments of wildlife communities in one of the world's most complex and biodiverse ecosystems.

AUTHOR CONTRIBUTIONS

Andrea Zampetti, Ana Benítez-López and Luca Santini conceived the idea. Andrea Zampetti, Ana Benítez-López and Iago Ferreiro-Arias collected the field data for the TROPECOLNET project. Andrea Zampetti, Iago Ferreiro-Arias, Laura Paltrinieri, Iván Ortiz and Brayan A. Cedeño-Panchez analysed the camera-trapping data from the TROPECOLNET project. Christophe Baltzinger, Christopher Beirne, Mark Bowler, Pierre-Michel Forget, Erik Guilbert, Yvonne Kemp, Carlos Peres, Andressa Scabin and Andrew Whitworth provided camera-trapping datasets for model training. Andrea Zampetti designed the methodology, performed data analyses and model training and led the writing of the manuscript. All authors contributed critically to the drafts and gave final approval for publication.

ACKNOWLEDGEMENTS

This work was supported by the TROPECOLNET project (ref: PID2022-138272NA-I00) funded by MCIN/AEI/10.13039/501100011033/FEDER, EU. Andrea Zampetti and Ana Benítez-López acknowledge support from the grant RYC2021-031737-I, funded by MCIN/AEI/10.13039/501100011033 and the EU ('NextGenerationEU'/PRTR). Andrea Zampetti, Ana Benítez-López and Iago Ferreiro-Arias acknowledge the Instituto Juruá and all the people involved, for their help and logistic support during fieldwork. All the authors acknowledge the multitude of people that contributed to fieldwork, data gathering and data processing for all the camera-trapping datasets used here. Andrea Zampetti also acknowledges Peter van Lunteren from AddaxAI for the critical role in making the AI algorithm publicly available and accessible.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/2041-210x.70213.

DATA AVAILABILITY STATEMENT

The TropiCam-AI algorithm is available and ready to use on the no-code AddaxAI platform (https://addaxdatascience.com/addaxai/). Further information on model usage can be found on the GitHub repository at https://github.com/andrewzamp/TropiCam-AI. All the codes used for model training and fine-tuning, along with the fully trained algorithm, iNaturalist images metadata and the full camera-trap dataset are also provided in a Zenodo repository at https://doi.org/10.5281/zenodo.17589042 (Zampetti et al., 2025).

ORCID

Andrea Zampetti https://orcid.org/0009-0001-0495-5416
Iago Ferreiro-Arias https://orcid.org/0000-0003-4178-5783
Mark T. Bowler https://orcid.org/0000-0001-5236-3477
Pierre-Michel Forget https://orcid.org/0000-0002-9252-974X
Andressa B. Scabin https://orcid.org/0000-0001-5377-2133

REFERENCES

Ahumada, J. A., Fegraus, E., Birch, T., Flores, N., Kays, R., O'Brien, T. G., Palmer, J., Schuttler, S., Zhao, J. Y., Jetz, W., Kinnaird, M., Kulkarni, S., Lyet, A., Thau, D., Duong, M., Oliver, R., & Dancer, A. (2020). Wildlife insights: A platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. *Environmental Conservation*, *47*(1), 1–6. https://doi.org/10.1017/S0376892919000298

Ahumada, J. A., Silva, C. E. F., Gajapersad, K., Hallam, C., Hurtado, J., Martin, E., McWilliam, A., Mugerwa, B., O'Brien, T., Rovero, F., Sheil, D., Spironello, W. R., Winarni, N., & Andelman, S. J. (2011). Community structure and diversity of tropical forest mammals: Data from a global camera trap network. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, *366*(1578), 2703–2711. https://doi.org/10.1098/rstb.2011.0115

Bánki, O., Roskov, Y., Döring, M., Ower, G., Hernández Robles, D. R., Plata Corredor, C. A., Stjernegaard Jeppesen, T., Örn, A., Pape, T., Hobern, D., Garnett, S., Little, H., DeWalt, R. E., Miller, J., Orrell, T., Aalbu, R., Abbott, J., Aedo, C., Aescht, E., … Şentürk, O. (2025). *Catalogue of life (Version 2025-07-10)*. Catalogue of Life Foundation. https://doi.org/10.48580/dgrrq

Beery, S., Agarwal, A., Cole, E., & Birodkar, V. (2021). The iWildCam 2021 competition dataset (No. arXiv:2105.03494). *arXiv*. https://doi.org/10.48550/arXiv.2105.03494

Beery, S., Morris, D., & Yang, S. (2019). Efficient pipeline for camera trap image review (No. arXiv:1907.06772). *arXiv*. https://doi.org/10.48550/arXiv.1907.06772

Beery, S. M., Cole, E., & Gjoka, A. (2020). The iWildCam 2020 competition dataset. *arXiv*. https://www.semanticscholar.org/paper/431717afcb02e83a1c77c08639199436d702dc4b

Bowler, M. T., Tobler, M. W., Endress, B. A., Gilmore, M. P., & Anderson, M. J. (2017). Estimating mammalian species richness and occupancy in tropical forest canopies with arboreal camera traps. *Remote Sensing in Ecology and Conservation*, *3*(3), 146–157. https://doi.org/10.1002/rse2.35

Brook, B. W., Buettel, J. C., van Lunteren, P., Rajmohan, P. P., & Aandahl, R. Z. (2025). MEWC: A user-friendly AI workflow for customised wildlife-image classification. *Peer Community Journal*, *5*, e57. https://doi.org/10.24072/pcjournal.565

Caravaggi, A., Banks, P. B., Burton, A. C., Finlay, C. M. V., Haswell, P. M., Hayward, M. W., Rowcliffe, M. J., & Wood, M. D. (2017). A review of camera trapping for conservation behaviour research. *Remote Sensing in Ecology and Conservation*, *3*(3), 109–122. https://doi.org/10.1002/RSE2.48

Coutant, O., Boissier, O., Ducrettet, M., Albert-Daviaud, A., Bouiges, A., Dracxler, C. M., Feer, F., Mendoza, I., Guilbert, E., & Forget, P.-M. (2022). Roads disrupt frugivory and seed removal in tropical animal-dispersed plants in French Guiana. *Frontiers in Ecology and Evolution*, *10*, 805376. https://doi.org/10.3389/fevo.2022.805376

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples (No. arXiv:1901.05555). *arXiv*. https://doi.org/10.48550/arXiv.1901.05555

Dussert, G., Chamaillé-Jammes, S., Dray, S., & Miele, V. (2025). Being confident in confidence scores: Calibration in deep learning models for camera trap image sequences. *Remote Sensing in Ecology and Conservation*, *11*(1), 88–99. https://doi.org/10.1002/rse2.412

Fennell, M., Beirne, C., & Burton, A. C. (2022). Use of object detection in camera trap image identification: Assessing a method to rapidly and accurately classify human and animal detections for research and application in recreation ecology. *Global Ecology and Conservation*, *35*, e02104. https://doi.org/10.1016/J.GECCO.2022.E02104

Gadot, T., Istrate, Ș., Kim, H., Morris, D., Beery, S., Birch, T., & Ahumada, J. (2024). To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Computer Vision*, *18*(8), 1193–1208. https://doi.org/10.1049/cvi2.12318

Gregory, T., Carrasco Rueda, F., Deichmann, J., Kolowski, J., & Alonso, A. (2014). Arboreal camera trapping: Taking a proven method to new heights. *Methods in Ecology and Evolution*, *5*(5), 443–451. https://doi.org/10.1111/2041-210X.12177

He, K., Zhang, X., Ren, S., & Sun, J. (2016). *Deep residual learning for image recognition*. https://openaccess.thecvf.com/content_cvpr_2016/html/He_Deep_Residual_Learning_CVPR_2016_paper.html

Hernandez, A., Miao, Z., Vargas, L., Beery, S., Dodhia, R., Arbelaez, P., & Ferres, J. M. L. (2024). Pytorch-wildlife: A collaborative deep learning framework for conservation (No. arXiv:2405.12930). *arXiv*. https://doi.org/10.48550/arXiv.2405.12930

Honda, A., Beirne, C., Huarcaya, R. P., Mullisaca, F. P., Quispe, C. Q., Quispe, R. V., Flatt, E., & Whitworth, A. (2025). Arboreal camera trapping reveals diel-vertical migrations in arboreal wildlife of the Peruvian Amazon rainforest. *Environmental Research: Ecology*, *4*(2), 025003. https://doi.org/10.1088/2752-664X/adb850

Howe, H. F., & Smallwood, J. (1982). Ecology of seed dispersal. *Annual Review of Ecology, Evolution, and Systematics*, *13*(1), 201–228. https://doi.org/10.1146/annurev.es.13.110182.001221

Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, *2015*(1), 258619. https://doi.org/10.1155/2015/258619

iNaturalist. (2025). Available from https://www.inaturalist.org. Accessed May, 2024.

Jenkins, C. N., Pimm, S. L., & Joppa, L. N. (2013). Global patterns of terrestrial vertebrate diversity and conservation. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(28), E2602–E2610. https://doi.org/10.1073/pnas.1302251110

Khan, K., Rehman, S. U., Aziz, K., Fong, S., & Sarasvady, S. (2014). *DBSCAN: Past, present and future*. The Fifth International Conference on the Applications of Digital Information and Web Technologies (ICADIWT 2014), 232–238. https://doi.org/10.1109/ICADIWT.2014.6814687

Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2018). Focal loss for dense object detection. *arXiv*, No. arXiv:1708.02002. https://doi.org/10.48550/arXiv.1708.02002

Link, A., & Di Fiore, A. (2006). Seed dispersal by spider monkeys and its importance in the maintenance of neotropical rain-forest diversity. *Journal of Tropical Ecology*, *22*, 235–246. https://doi.org/10.1017/S0266467405003081

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). *A ConvNet for the 2020s*. https://openaccess.thecvf.com/content/CVPR2022/html/Liu_A_ConvNet_for_the_2020s_CVPR_2022_paper.html

MacKenzie, D. I., Nichols, J. D., Royle, J. A., Pollock, K. H., Bailey, L., & Hines, J. E. (2017). *Occupancy estimation and modeling: Inferring patterns and dynamics of species occurrence*. Elsevier.

Mitterwallner, V., Peters, A., Edelhoff, H., Mathes, G., Nguyen, H., Peters, W., Heurich, M., & Steinbauer, M. J. (2024). Automated visitor and wildlife monitoring with camera traps and machine learning. *Remote Sensing in Ecology and Conservation*, *10*(2), 236–247. https://doi.org/10.1002/rse2.367

Moore, J. F., Soanes, K., Balbuena, D., Beirne, C., Bowler, M., Carrasco-Rueda, F., Cheyne, S. M., Coutant, O., Forget, P.-M., Haysom, J. K., Houlihan, P. R., Olson, E. R., Lindshield, S., Martin, J., Tobler, M., Whitworth, A., & Gregory, T. (2021). The potential and practice of arboreal camera trapping. *Methods in Ecology and Evolution*, *12*(10), 1768–1779. https://doi.org/10.1111/2041-210X.13666

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Nathan, R., & Muller-Landau, H. C. (2000). Spatial patterns of seed dispersal, their determinants and consequences for recruitment. *Trends in Ecology & Evolution*, *15*(7), 278–285. https://doi.org/10.1016/S0169-5347(00)01874-7

Norouzzadeh, M. S., Morris, D., Beery, S., Joshi, N., Jojic, N., & Clune, J. (2021). A deep active learning system for species identification and counting in camera trap images. *Methods in Ecology and Evolution*, *12*(1), 150–161. https://doi.org/10.1111/2041-210X.13504

Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115

Peral, C., Landman, M., & Kerley, G. I. H. (2022). The inappropriate use of time-to-independence biases estimates of activity patterns of free-ranging mammals derived from camera traps. *Ecology and Evolution*, *12*(10), e9408. https://doi.org/10.1002/ece3.9408

Peres, C. A., & van Roosmalen, M. (2002). Patterns of primate frugivory in Amazonia and the Guianan shield: Implications to the demography of large-seeded plants in overhunted tropical forests. In D. J. Levey, W. R. Silva, & M. Galetti (Eds.), *Seed dispersal and frugivory: Ecology, evolution and conservation* (pp. 407–423). CABI International.

Pigot, A. L., Sheard, C., Miller, E. T., Bregman, T. P., Freeman, B. G., Roll, U., Seddon, N., Trisos, C. H., Weeks, B. C., & Tobias, J. A. (2020). Macroevolutionary convergence connects morphological form to ecological function in birds. *Nature Ecology & Evolution*, *4*(2), 230–239. https://doi.org/10.1038/s41559-019-1070-4

Pillay, R., Venter, M., Aragon-Osejo, J., González-del-Pliego, P., Hansen, A. J., Watson, J. E., & Venter, O. (2022). Tropical forests are home to over half of the world's vertebrate species. *Frontiers in Ecology and the Environment*, *20*(1), 10–15. https://doi.org/10.1002/fee.2420

Rigoudy, N., Dussert, G., Benyoub, A., Besnard, A., Birck, C., Boyer, J., Bollet, Y., Bunz, Y., Caussimont, G., Chetouane, E., Carriburu, J. C., Cornette, P., Delestrade, A., De Backer, N., Dispan, L., Le Barh, M., Duhayer, J., Elder, J.-F., Fanjul, J.-B., … Chamaillé-Jammes, S. (2023). The DeepFaune initiative: A collaborative effort towards the automatic identification of European fauna in camera trap images. *European Journal of Wildlife Research*, *69*(6), 113. https://doi.org/10.1007/s10344-023-01742-7

Scabin, A. B., & Peres, C. A. (2021). Hunting pressure modulates the composition and size structure of terrestrial and arboreal vertebrates in Amazonian forests. *Biodiversity and Conservation*, *30*(12), 3613–3632. https://doi.org/10.1007/s10531-021-02266-9

Schneider, S., Greenberg, S., Taylor, G. W., & Kremer, S. C. (2020). Three critical factors affecting automated image species recognition performance for camera traps. *Ecology and Evolution*, *10*(7), 3503–3517. https://doi.org/10.1002/ece3.6147

Séguigne, M., Coutant, O., Bouton, B., Picart, L., Guilbert, É., & Forget, P.-M. (2022). Arboreal camera trap reveals the frequent occurrence of a frugivore-carnivore in neotropical nutmeg trees. *Scientific Reports*, *12*(1), 7513. https://doi.org/10.1038/s41598-022-11568-z

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Shepley, A., Falzon, G., Meek, P., & Kwan, P. (2021). Automated location invariant animal detection in camera trap images using publicly available data sources. *Ecology and Evolution*, *11*(9), 4494–4506. https://doi.org/10.1002/ece3.7344

Silveira, L., Jácomo, A. T. A., & Diniz-Filho, J. A. F. (2003). Camera trap, line transect census and track surveys: A comparative evaluation. *Biological Conservation*, *114*(3), 351–355. https://doi.org/10.1016/S0006-3207(03)00063-6

Simonyan, K. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv Preprint arXiv:1409.1556*. https://gitea.sharpe6.com/Adog64/Adversarial-Machine-Learning-Clinic/raw/commit/0ba4e489bb77a0d2642923499598e309d8751435/references/Very_Deep_Convolutional_Networks_for_Large-Scale_Image_Recognition.pdf

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* (Vol. 25). https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

Stevens, S., Wu, J., Thompson, M. J., Campolongo, E. G., Song, C. H., Carlyn, D. E., Dong, L., Dahdul, W. M., Stewart, C., Berger-Wolf, T., Chao, W.-L., & Su, Y. (2024). *BioCLIP: A vision foundation model for the tree of life*. https://openaccess.thecvf.com/content/CVPR2024/html/Stevens_BioCLIP_A_Vision_Foundation_Model_for_the_Tree_of_Life_CVPR_2024_paper.html

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Newton, E. J., Boughton, R. K., Ivan, J. S., Odell, E. A., Newkirk, E. S., Conrey, R. Y., Stenglein, J., Iannarilli, F., Erb, J., Brook, R. K., Davis, A. J., Lewis, J., Walsh, D. P., Beasley, J. C., VerCauteren, K. C., Clune, J., & Miller, R. S. (2020). Improving the accessibility and transferability of machine learning algorithms for identification of animals in camera trap images: MLWIC2. *Ecology and Evolution*, *10*(19), 10374–10383. https://doi.org/10.1002/ece3.6692

Tabak, M. A., Norouzzadeh, M. S., Wolfson, D. W., Sweeney, S. J., Vercauteren, K. C., Snow, N. P., Halseth, J. M., Di Salvo, P. A., Lewis, J. S., White, M. D., Teton, B., Beasley, J. C., Schlichting, P. E., Boughton, R. K., Wight, B., Newkirk, E. S., Ivan, J. S., Odell, E. A., Brook, R. K., … Miller, R. S. (2019). Machine learning to classify animal species in camera trap images: Applications in ecology. *Methods in Ecology and Evolution*, *10*(4), 585–590. https://doi.org/10.1111/2041-210X.13120

Tilman, D., Clark, M., Williams, D. R., Kimmel, K., Polasky, S., & Packer, C. (2017). Future threats to biodiversity and pathways to their prevention. *Nature*, *546*(7656), 73–81. https://doi.org/10.1038/nature22900

Tobler, M. W., Zúñiga Hartley, A., Carrillo-Percastegui, S. E., & Powell, G. V. N. (2015). Spatiotemporal hierarchical modelling of species richness and occupancy using camera trap data. *Journal of Applied Ecology*, *52*(2), 413–421. https://doi.org/10.1111/1365-2664.12399

Tuba, E., Bačanin, N., Strumberger, I., & Tuba, M. (2021). Convolutional neural networks Hyperparameters tuning. In E. Pap (Ed.), *Artificial intelligence: Theory and applications* (pp. 65–84). Springer International Publishing. https://doi.org/10.1007/978-3-030-72711-6_4

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., Adam, H., Perona, P., & Belongie, S. (2018). *The INaturalist species classification and detection dataset*. https://openaccess.thecvf.com/content_cvpr_2018/html/Van_Horn_The_INaturalist_Species_CVPR_2018_paper.html

Van Lunteren, P. (2023). EcoAssist: A no-code platform to train and deploycustom YOLOv5 object detection models. *Journal of Open Source Software*, *8*(88), 5581. https://doi.org/10.21105/joss.05581

Vendrow, E., Pantazis, O., Shepard, A., Brostow, G., Jones, K. E., Aodha, O. M., Beery, S., & Horn, G. V. (2024). INQUIRE: A natural world text-to-image retrieval benchmark (No. arXiv:2411.02537). *arXiv*. https://doi.org/10.48550/arXiv.2411.02537

Villalva, P., & Jordano, P. (2025). A machine learning application to camera-traps: Robust species interactions datasets for analysis of mutualistic networks. *bioRxiv*. https://doi.org/10.1101/2025.03.01.640990

Whitworth, A., Beirne, C., Pillco Huarcaya, R., Whittaker, L., Serrano Rojas, S. J., Tobler, M. W., & MacLeod, R. (2019). Human disturbance impacts on rainforest mammals are most notable in the canopy, especially for larger-bodied species. *Diversity and Distributions*, *25*(7), 1166–1178. https://doi.org/10.1111/ddi.12930

Whitworth, A., Braunholtz, L. D., Huarcaya, R. P., MacLeod, R., & Beirne, C. (2016). Out on a limb: Arboreal camera traps as an emerging methodology for inventorying elusive rainforest mammals. *Tropical Conservation Science*, *9*(2), 675–698. https://doi.org/10.1177/194008291600900208

Whytock, R. C., Świeżewski, J., Zwerts, J. A., Bara-Słupski, T., Koumba Pambo, A. F., Rogala, M., Bahaa-el-din, L., Boekee, K., Brittain, S., Cardoso, A. W., Henschel, P., Lehmann, D., Momboua, B., Kiebou Opepa, C., Orbell, C., Pitman, R. T., Robinson, H. S., & Abernethy, K. A. (2021). Robust ecological analysis of camera trap data labelled by a machine learning model. *Methods in Ecology and Evolution*, *12*(6), 1080–1092. https://doi.org/10.1111/2041-210X.13576

Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., & Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, *10*(1), 80–91. https://doi.org/10.1111/2041-210X.13099

Zampetti, A., Mirante, D., Palencia, P., & Santini, L. (2024). Towards an automated protocol for wildlife density estimation using camera-traps. *Methods in Ecology and Evolution*, *15*(12), 2276–2288. https://doi.org/10.1111/2041-210X.14450

Zampetti, A., Santini, L., Ferreiro Arias, I., Paltrinieri, L., Ortiz, I., Cedeño-Panchez, B., Baltzinger, C., Beirne, C., Bowler, M., Forget, P.-M., Guilbert, E., Kemp, Y., Peres, C., Scabin, A. B., Whitworth, A., & Benítez López, A. (2025). Data, code and model for: "introducing TropiCam-AI: A taxonomically flexible automated classifier of Neotropical arboreal mammals and birds from camera-trap data". In Methods in ecology and evolution. *Zenodo*. https://doi.org/10.5281/zenodo.17589043

Zhu, C., Li, W., Gregory, T., Wang, D., Ren, P., Zeng, D., Kang, Y., Ding, P., & Si, X. (2022). Arboreal camera trapping: A reliable tool to monitor plant-frugivore interactions in the trees on large scales. *Remote Sensing in Ecology and Conservation*, *8*(1), 92–104. https://doi.org/10.1002/rse2.232

SUPPORTING INFORMATION

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

Figure S1. Training and validation accuracy curves over the 50 epochs of training.

Figure S2. Training and validation accuracy losses, measured via cross entropy, over the 50 epochs of training.

Figure S3. Overall model recall for different thresholds of taxonomic aggregation, where the colour gradient indicates the proportion of correct prediction at species level for that threshold value.

Figure S4. Confusion matrix for predictions at species level on the testing dataset.

Figure S5. Confusion matrix for predictions at genus level on the testing dataset.

Figure S6. Confusion matrix for predictions at family level on the testing dataset.

Figure S7. Confusion matrix for predictions at order level on the testing dataset.

Table S1. Number of images used in the training, validation and testing phase to develop TropiCam-AI.

Table S2. Selected hyperparameters to fine-tune in the Bayesian optimization phase.

Table S3. Recall parameters for all taxa included in the model, computed on the testing set.

Table S4. Performance metrics of the model's predictions on the out-of-sample Brazilian dataset.

Supporting Information 2. Confusion matrix showing predicted versus true taxonomic classifications at the species level.

How to cite this article: Zampetti, A., Santini, L., Ferreiro-Arias, I., Paltrinieri, L., Ortiz, I., Cedeño-Panchez, B. A., Baltzinger, C., Beirne, C., Bowler, M. T., Forget, P.-M., Guilbert, E., Kemp, Y. J. M., Peres, C. A., Scabin, A. B., Whitworth, A., & Benítez-López, A. (2026). Introducing TropiCam-AI: A taxonomically flexible automated classifier of Neotropical arboreal mammals and birds from camera-trap data. *Methods in Ecology and Evolution*, *00*, 1–13. https://doi.org/10.1111/2041-210x.70213