Muir, Jen, Ravuri, Aditya, Meissner, Eric, Hawes, Joseph ORCID: https://orcid.org/0000-0003-0053-2018 , Dufourq, Emmanuel ORCID: https://orcid.org/0000-0002-6986-3408 , O'Mahoney, Thomas and Dunn, Jacob (2025) Detection and classification of captive coppery titi monkey calls. Bioacoustics .

Downloaded from: https://insight.cumbria.ac.uk/id/eprint/8860/

Usage of any items from the University of Cumbria's institutional repository 'Insight' must conform to the following fair usage guidelines.

Any item and its associated metadata held in the University of Cumbria's institutional repository Insight (unless stated otherwise on the metadata record) may be copied, displayed or performed, and stored in line with the JISC fair dealing guidelines (available here) for educational and not-for-profit activities

provided that

- the authors, title and full bibliographic details of the item are cited clearly when any part of the work is referred to verbally or in the written form
- a hyperlink/URL to the original Insight record of that item is included in any citations of the work
- the content is not changed in any way
- all files required for usage of the item are kept together with the main item file.

You may not

- sell any part of an item
- refer to any part of an item without citation
- amend any item or contextualise it in a way that will impugn the creator's reputation
- remove or alter the copyright statement on an item.

The full policy can be found here.
Alternatively contact the University of Cumbria Repository Editor by emailing insight@cumbria.ac.uk.

Bioacoustics

The International Journal of Animal Sound and its Recording

ISSN: 0952-4622 (Print) 2165-0586 (Online) Journal homepage: www.tandfonline.com/journals/tbio20

Detection and classification of captive coppery titi monkey calls

Jen Muir, Aditya Ravuri, Eric Meissner, Joseph E. Hawes, Emmanuel Dufourq, Thomas O'Mahoney & Jacob C. Dunn

To cite this article: Jen Muir, Aditya Ravuri, Eric Meissner, Joseph E. Hawes, Emmanuel Dufourq, Thomas O'Mahoney & Jacob C. Dunn (22 May 2025): Detection and classification of captive coppery titi monkey calls, Bioacoustics, DOI: 10.1080/09524622.2025.2496497

To link to this article: https://doi.org/10.1080/09524622.2025.2496497

BIOACOUSTICS
https://doi.org/10.1080/09524622.2025.2496497

Taylor & Francis
Taylor & Francis Group

Detection and classification of captive coppery titi monkey calls

Jen Muir[a], Aditya Ravuri[b], Eric Meissner[b], Joseph E. Hawes[c], Emmanuel Dufourq[d], Thomas O'Mahoney[a] and Jacob C. Dunn[a,e,f*]

[a]School of Life Sciences, Anglia Ruskin University, Cambridge, UK; [b]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK; [c]Institute of Science and Environment, University of Cumbria, Ambleside, UK; [d]African Institute for Mathematical Sciences - Research and Innovation Centre, Kigali, RW, South Africa; [e]Biological Anthropology, University of Cambridge, Cambridge, UK; [f]Department of Behavioural and Cognitive Biology, University of Vienna, Vienna, Austria

ABSTRACT
Ecoacoustic monitoring has many applications in conservation and welfare but generates large amounts of data that are extremely time-intensive to manually process. This has led to an increased interest in the use of machine learning methods to increase efficiency and reduce workload. Common issues within this area include noisy, unbalanced and limited datasets, making it challenging to make effective machine learning models. This study aimed to determine the vocal repertoire of the coppery titi monkey, *Plecturocebus cupreus*, and develop a machine learning model that can detect, segment and classify calls within streaming audio using a small and unbalanced dataset with overlapping calls from other species. Acoustic data were collected across three zoo populations of *P. cupreus* using passive acoustic monitors. From this, 3302 calls were manually labelled to use as training data. Ten call types were established manually, corresponding to three groups: short calls, long calls and harsh calls. A Long Short-Term Memory neural network was created that successfully detected calls (accuracy = 0.95) and classified call types (accuracy = 0.97). Potential applications for the model include welfare monitoring in captivity and population monitoring of *P. cupreus* and related endangered species in the wild.

ARTICLE HISTORY
Received 21 October 2024
Accepted 16 April 2025

KEYWORDS
Vocal repertoire; classification; detection; machine learning; primate communication; Pitheciidae

Introduction

Bioacoustic research and ecoacoustic monitoring often generate large quantities of data for which the manual scanning and labelling of sounds of interest is time-consuming and often requires expert knowledge. This time-intensive process is often impractical for the datasets generated by passive acoustic monitoring (PAM) devices, which may include many months' worth of audio data, often summing multiple terabytes. Given this, Machine Learning (ML) approaches are increasingly being used to automatically detect and classify

---

CONTACT Jen Muir jenmuir93@gmail.com

*Current affiliation: ENES Bioacoustics Research Laboratory, University of Saint .tienne, St-.tienne 42023, France

Supplemental data for this article can be accessed online at https://doi.org/10.1080/09524622.2025.2496497.

sounds of interest in order to mitigate the issues of processing such large datasets (Mcloughlin et al. 2019; Stowell 2022; Nieto-Mora et al. 2023; Kershenbaum et al. 2025).

ML encompasses a wide variety of techniques that aim to use input data to make predictions, each varying in complexity and suitability for a given task (Jo 2021). Algorithms may implement a supervised, semi-supervised or unsupervised approach. Supervised approaches are most common within bioacoustic research and require manually labelled data to train models (Bianco et al. 2019). These models can range from simple logistic regression (Tseng et al. 2021; Townsend and Aldstadt 2023) to more complex support vector machines (SVM; Noda et al. 2016; Chao et al. 2019; Clink et al. 2019), and deep learning methods such as neural networks (NN; Bermant 2021; Gong et al. 2021; Morales et al. 2022), convolutional neural networks (CNN; Merchan et al. 2020; Xu et al. 2020; Paumen et al. 2022) and recurrent neural networks (RNN; Cakir et al. 2017; Wang et al. 2022; Srujana et al. 2023). Model success can be assessed using metrics such as accuracy, precision, F-score, and recall, all of which are common within supervised approaches (Bianco et al. 2019). Unsupervised approaches are used to determine patterns within unlabelled data by grouping items through algorithms such as Gaussian Mixture Models (GMM; Diep et al. 2013; Parada and Cardenal-López 2014; Zhao et al. 2017), autoencoders (Ibrahim et al. 2019; Alipek et al. 2023; Vamsi et al. 2023), hidden Markov models (HMM; Trawicki 2021; Ogundile et al. 2022), and hierarchical or k-means clustering (Ozanich et al. 2021; Batist et al. 2023; Jiang et al. 2024). Results can then be visualised using dimensionality reduction methods such as t-SNE (t-distributed stochastic neighbour embedding; Van der Maaten and Hinton 2008) or UMAP (uniform manifold approximation and projection; McInnes et al. 2018). Dimensionality reduction may be used alone or in combination with supervised methods. However, the lack of human input makes these methods difficult to assess in terms of efficacy, and so they are often used to initially identify patterns of interest in data before further analysis (e.g. to suggest call groupings within a vocal repertoire; Valente et al. 2019; Goffinet et al. 2021; Thomas et al. 2022). Semi-supervised approaches exist between supervised and unsupervised, using labelled data alongside unlabelled data for training a model. These models may use similar architectures (i.e. model structure) to supervised models and be assessed using methods designed for supervised or unsupervised models (e.g. F1 score, dimensionality reduction).

Common tasks that ML has been applied to within bioacoustics include the detection, segmentation and classification of animal calls (Stowell 2022; Oswald et al. 2022; Kershenbaum et al. 2025). Detection tasks simply aim to determine if a call of interest is present within the data, and it is often the first step before segmentation (a process in which the signal is isolated from the rest of the recording, usually based on given start/end times or amplitude thresholds) or classification of the target call (e.g. Coffey et al. 2019; Garcia et al. 2020; Zhong et al. 2021). Detection algorithms have been applied within animal conservation to identify the calls of target species and have been particularly effective for presence-absence surveys of cryptic species that would be difficult to detect otherwise (e.g. bats: Mac Aodha et al. 2018; least bittern, *Ixobrychus exilis*:; Znidersic et al. 2020; Kroombit tinker frog, *Taudactylus pleione*:; Xie et al. 2022). Others have searched for rare acoustic events, such as gunshots associated with poaching (Katsis et al. 2022; Somwong et al. 2023), wildfires (Huang et al. 2022; Grari et al. 2023), and illegal logging (Srisuphab et al. 2020). Signal detection has also been used for studying animal welfare from calls, including the detection of stress-related calls (Schön et al. 2001Nasirahmadi et al. 2020; Mao et al. 2022), calls of positive valence (Briefer et al. 2022), and calls that correlate with illness (Carroll et al. 2014; Shen et al. 2022) or oestrus (Chung et al. 2013). In many cases, the detected signal is then automatically segmented. Automated call segmentation has been used to isolate syllables of frog calls (Xie et al. 2020), cetacean calls (Schröter et al. 2019) and birdsong (Koops et al. 2015; Narasimhan et al. 2017; Zsebők et al. 2021), as well as to annotate data (e.g. labelling call types or components; Steinfath et al. 2021). Classification models are designed to take segmented data and categorise them based on a predetermined list of labels. This may include species (Zualkernan et al. 2021; Akbal et al. 2022; Noumida and Rajan 2022), caller identity (Larranaga et al. 2015; Trapanotto et al. 2022; Phaniraj 2022), or call type (Oikarinen et al. 2019; Jung et al. 2021; Bergler et al. 2022).

While the use of ML can be very effective for reducing workloads and improving accuracy in processing bioacoustic data, there are still difficulties present in designing ML solutions. Data from field recordings can include background noise, making it difficult to isolate the calls of interest, which may also overlap with calls from other individuals or species (Arnaud et al. 2023). Call types within a species' repertoire are also unlikely to be evenly emitted, with some call types being much more common than others, leading to an unbalanced dataset than can be difficult to train ML models on. Issues can also arise from the graded nature of some call repertoires (e.g. pigs: Tallet et al. 2013; baboons:; Wadewitz et al. 2015; rooks:; Martin et al. 2024), with calls existing on a continuum rather than as discrete categories, which can reduce accuracy when classifying call types. There are also limited data available on the vocalisations of many species, making the development of detection models problematic (Arnaud et al. 2023). These challenges mean that ML models need to be effective when working with small, unbalanced, and noisy datasets.

A good case study for testing ML models from bioacoustic datasets is the titi monkeys (Pitheciidae: Callicebinae), a highly vocal South American primate group for which there is an existing body of research concerning their vocalisations (Bezerra et al. 2017). Vocal repertoires have been described for the red-bellied titi monkey, *Plecturocebus moloch*, (Moynihan 1966; Robinson 1979), the collared titi monkey, *Cheracebus torquatus* (Kinzey et al. 1977), and the black-fronted titi monkey, *Callicebus nigrifrons* (Cäsar 2012; Cäsar et al. 2012), detailing between six and 13 call types. Van Kuijk et al. (2023) have also successfully developed an ML approach to detect red titi monkey, *Plecturocebus discolor*, duets in the wild. The vocalisations of this group are ideal for testing models on challenging datasets due to the presence of noisy and graded calls in their repertoire, and unbalanced frequency of call use (Moynihan 1966; Robinson 1979). As a vocal species, this also enables large amounts of data to be gathered quickly.

Around 32% of titi monkey species are threatened in the wild, primarily due to habitat loss and fragmentation (IUCN 2024). Due to their cryptic nature and hard-to-access habitat, researching and monitoring their populations can be challenging (Pinto et al. 2013; Souza-Alves et al. 2023). As titi monkey loud calls are species distinct, ML models for call detection could be ideal for population monitoring in the wild of at-risk titi monkey species that would otherwise be difficult to find (Aldrich et al. 2008). Titi monkeys can also be difficult to keep in captivity, as they are sensitive to stress (Mendoza and Mason 1997), experience high infant mortality (Veiga et al. 2013), and are prone to respiratory conditions (Jones 1968; Lorenz and Mason 1971). Monitoring in captivity could present a non-invasive way to assess health and welfare, as well as to learn more about what may provoke stress in these species and to inform care. To achieve this an understanding of the behavioural contexts and associated stress levels accompanying each call type would be required to fully develop a welfare monitoring system. A successful detection and classification model could thus be usefully applied within conservation and welfare contexts.

This study focused on the vocalisations of captive coppery titi monkeys, *Plecturocebus cupreus*, within UK zoos. This species was chosen due to the challenges of housing them within zoos, making it a species of interest for welfare monitoring in captivity. Specifically, we aimed to: (i) catalogue the vocal repertoire for *P. cupreus* in captivity; (ii) develop a ML pipeline to detect and segment *P. cupreus* calls in streaming audio; and (iii) classify calls according to the vocal repertoire.

Methods

Study population

A total of 20 individuals across five groups were sampled from the three zoos. Shaldon Zoo had one group composed of three individuals (one adult male, two adult females), Banham Zoo housed a single group of five individuals (two adult males, three adult females), and Blackpool Zoo housed three groups, two with three individuals (one adult male, two adult females; three adult males) and one with five (two adult males, three adult females). The use of data from five separate groups was chosen to increase the generalisability of the sample, reducing the likelihood of certain individuals contributing excessively.

Audio data collection

Audio data were collected via passive acoustic monitoring at Banham Zoo, Blackpool Zoo, and Shaldon Zoo between July 2021 and January 2023. AudioMoths (Hill et al. 2018) were placed within indoor enclosures and set to record continuously with a sampling rate of 44.1 kHz from 30 minutes before sunrise to 30 minutes after sunset each day to capture the full titi monkey activity period. Recordings were made as a series of 10-minute files. A total of 100 days of data were collected; 15 days of data were collected from one group at Banham Zoo (July 2022), 10 days from one group at Shaldon Zoo (July 2021) and 25 days from each of three groups at Blackpool Zoo (December 2022–January 2023). This provided a total of 33,180 minutes of audio data.

Vocal repertoire

A vocal repertoire was created, comprising a comprehensive list of call types for the proposed model to detect and classify. Call type names were chosen in relation to previously published repertoires by Robinson (1979) and Moynihan (1966). The repertoire was first determined qualitatively using a subset of the data, consisting of calls exclusively from Banham Zoo ($N = 879$). A t-SNE (Van der Maaten and Hinton 2008) was then used to visually cluster calls into two-dimensional space. The repertoire was confirmed based on the clustering of call types found in the t-SNE. This method has previously been used as a tool for determining how structurally discrete calls are within a repertoire, and to visually assess how much call types are graded in terms of structure (e.g. overlaps in duration, frequency range, modulation between call types) (Valente et al. 2019; Goffinet et al. 2021; Thomas et al. 2022). The t-SNE was conducted with the scikit-learn library in Python (Pedregosa et al. 2011), using 20 MFCCs (Mel Frequency Cepstral Coefficients) and default parameters. Only calls from Banham Zoo were used to avoid the influence of differing soundscapes on the clustering.

To ensure call categorisation was consistent, inter-rater reliability was tested using Cohen's Kappa between two raters using 256 (30%) of the calls from the Banham dataset and split evenly between the call types (Landis and Koch 1977), with high agreement found between raters (K = 0.87, 95% CIs: 0.825–0.915, $p < 0.001$).

Training/testing data

From each zoo, 30 minutes of audio data were selected to use as training/testing data, corresponding to 90 minutes and 3,302 calls in total. Calls were defined as either target (i.e. calls produced by *P. cupreus*) or non-target (i.e. any sound produced by the environment or a different species). Non-target calls were also cross-referenced with data from the Macaulay Library (https://www.macaulaylibrary.org). These non-target calls were produced by a variety of species and at differing levels between zoos (e.g. Blackpool Zoo had very little noise from neighbouring species due to enclosure design, while Banham Zoo had many other species). Non-target species identified included Goeldi's marmosets, *Callimico goeldii*, sun conures, *Aratinga solstitialis*, common gulls, *Larus canus*, Eurasian blackbirds, *Turdus merula* and human voices, all of which overlapped with the frequency range of *P. cupreus* calls (Figure 1).

Each call in the training/testing dataset was manually labelled with its start/end time and call type in Praat (v 6.4.05; Boersma and Weenink 2024), with 3,302 calls overall

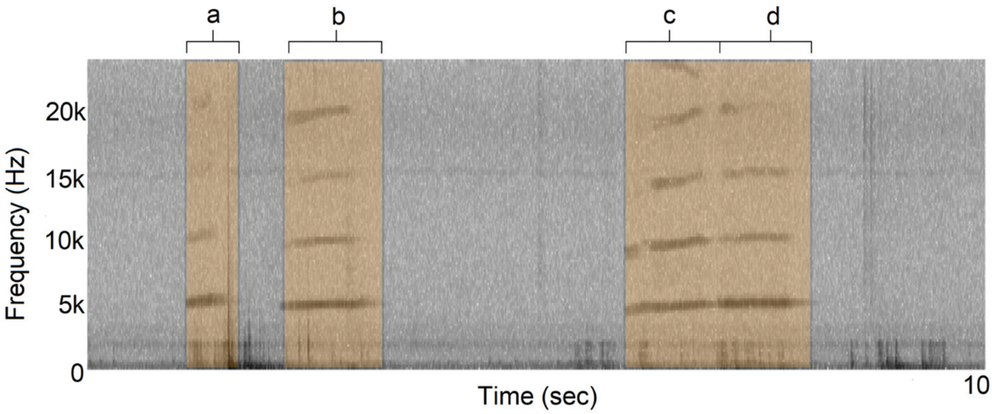

Figure 1. View from the Dash app (Hossain et al. 2019) for a sample across 10 seconds of acoustic data from captive coppery titi monkeys, *Plecturocebus cupreus*, showing detected chirp (a) and three whistles (b–d) in yellow.

labelled in the 90 mins of training data. As certain calls were found to be highly graded, a simplified 6-item version of the full vocal repertoire was used for model training and testing, combining the short and long call groups into two call categories: Chucks and Chirps were grouped as 'Short Calls' and Trills, Whistles, and Phees were grouped as 'Long Calls'. Jagged Trill and the Harsh Calls (Resonating Note, Sneeze, and Moan) were kept as independent call types. Scream calls were excluded from the ML pipeline due to a lack of examples ($n < 10$). When labelling training data, calls were labelled by these condensed categories as initial attempts to classify by each of the ten individual call types were unsuccessful. This is likely due to the spectral similarity of some call types (e.g. chirps and whistles share a similar frequency range and there is no clear duration distinction due to grading). Given this, our training data included Short Calls ($N = 1{,}264$), Long Calls ($N = 867$), Jagged Trills ($N = 96$), Resonating Notes ($N = 952$), Sneezes ($N = 23$) and Moans ($N = 100$). Occurrences of non-target vocalisations were also labelled ($N = 1{,}839$).

Once the training/testing calls were labelled, a randomly selected 90:10 (training:test) split was used for testing the detection model, while an 80:20 split was used to test the classification model (Sugali et al. 2024). Calls were stratified to ensure balance between call types when splitting the data.

Final model architecture

Detection

For call detection a Long Short-Term Memory network (LSTM) was developed using the Pytorch (Paszke et al. 2019), librosa (McFee et al. 2015) and SciPy (Virtanen et al. 2020) libraries in Python. The LSTM works as an enhanced version of an RNN, which can retain long-term information, which RNN models can struggle with (Hochreiter and Schmidhuber 1997; Gers et al. 2000).

Prior to running the model, the audio data were converted to a short-time Fourier transform (STFT) representation, transformed into a Mel-Frequency Cepstral Coefficient representation ($n = 40$ MFCCs), and finally normalised. The network had three stacked LSTM cells, each with 16 hidden units. This was followed by a sigmoid unit. The proposed call detection model takes in data of the shape of [552, 40], where 552 is the number of time steps in an audio clip and 40 corresponds to the MFCCs. For each input, the network produces 552 values, representing the probability of whether a call has been detected in each time step. LSTMs, a type of recurrent neural network, have successfully been applied in bioacoustics (Islam and Valles 2020; Duan et al. 2021; Wang et al. 2022).

Classification

To classify calls, we fit a similar model, with the only difference in the architecture being that the last layer of the neural network performed multiclass classification (where classes were the six call categories, plus an additional class for 'not call' as opposed to 'call' and 'not call'). This was achieved by thresholding the output values by 0.5 (i.e. <0.5 = absence class, presence class otherwise).

Both models were fit using the labelled training dataset using a standard cross entropy loss function (Zhang and Sabuncu 2018; Mao et al. 2023). A Dash app (a custom made interactive web application; Hossain et al. 2019) was constructed to assess the model's initial performance and to choose hyperparameters by visualising call segmentation and classification within spectrograms (Figure 1).

Model evaluation

Four performance metrics were used to evaluate the model detection and classification success: precision, recall, accuracy, and F1 score (Handelman et al. 2019). These metrics are derived from the false positives (fp), false negatives (fp), true positives (tp), and true negatives (tn), given in confusion matrices. Precision measures the proportion of correctly identified positive results out of all positive results, both true and false (Equation 1). This gives a measure of how often the model successfully predicts the target classes. Recall (also called sensitivity) measures the proportion of correctly identified positive results out of all positively labelled samples (Equation 2). Accuracy measures the proportion of correctly identified samples out of all samples (Equation 3). F1 score is defined as the harmonic mean of the precision and recall scores (Equation 4). Both precision and F1 score are recommended for unbalanced datasets, where accuracy alone can be misleading (Lever 2016). All measures are between 0 (low performance) and 1 (high performance). When determining at which score the model should be considered successful, 70% was chosen as a threshold. While there is no perfect threshold for success in machine learning, this was chosen as a balance between model functionally while considering the irregularities of acoustic data (Kershenbaum et al. 2025).

$$\text{Precision} = \frac{tp}{tp+fp} \tag{1}$$

$$\text{Recall} = \frac{tp}{tp+fn} \tag{2}$$

$$\text{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} \tag{3}$$

$$\text{F1 Score} = 2 \times \frac{precision \times recall}{precision + recall} \tag{4}$$

Data Availability

All audio data are available on Zenodo (https://doi.org/10.5281/zenodo.11040660). The code for the detection and classification models are available from: https://github.com/mlatcl/cmci/blob/main/src/main.py.

Results

Vocal repertoire

Ten call types in total were determined within the repertoire, corresponding to three groups; short calls, long calls, and harsh calls (Figure 3). These names were derived from previous repertoires of other titi monkey species (*P. moloch*; Moynihan 1966; Robinson, 1979):

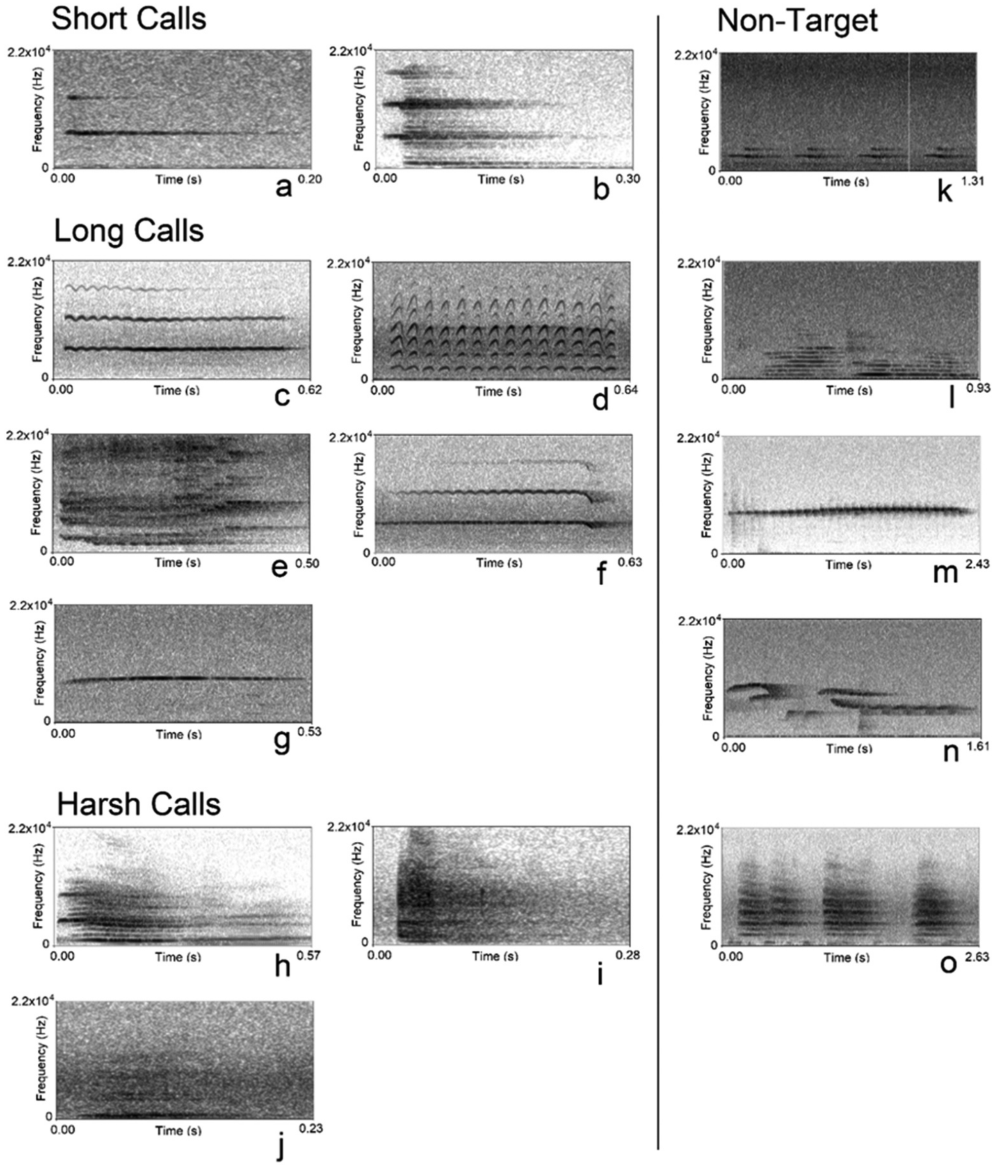

Figure 2. Spectrograms for call groups (Short calls, Long calls, Harsh calls) and types found in the vocal repertoire of captive coppery titi monkeys, *Plecturocebus cupreus*, and non-target calls from other species. Short calls: a) Chirp, b) Chuck; Long calls: c) Trill, d) Jagged Trill, e) Scream, f) Phee, g) Whistle; Harsh calls: h) Resonating Note, i) Sneeze, j) Moan; and Non-target calls: k) common gull, *Larus canus;* l) human speech; m) Goeldi's marmoset, *Callimico goeldii*; n) Eurasian blackbird, *Turdus merula*; and o) sun conure, *Aratinga solstitialis*.

Short calls include chirp and chuck calls and are short and modulated (Figure 2(a,b) Long calls are composed of trills, jagged trills, screams, whistles and phees (Figure 2(c-g). These calls are the longest in the repertoire and are often modulated, either increasing or decreasing slightly in frequency. Phee calls are typically modulated upwards at the start and become more broadband at the end. Harsh calls include moan, resonating note, and

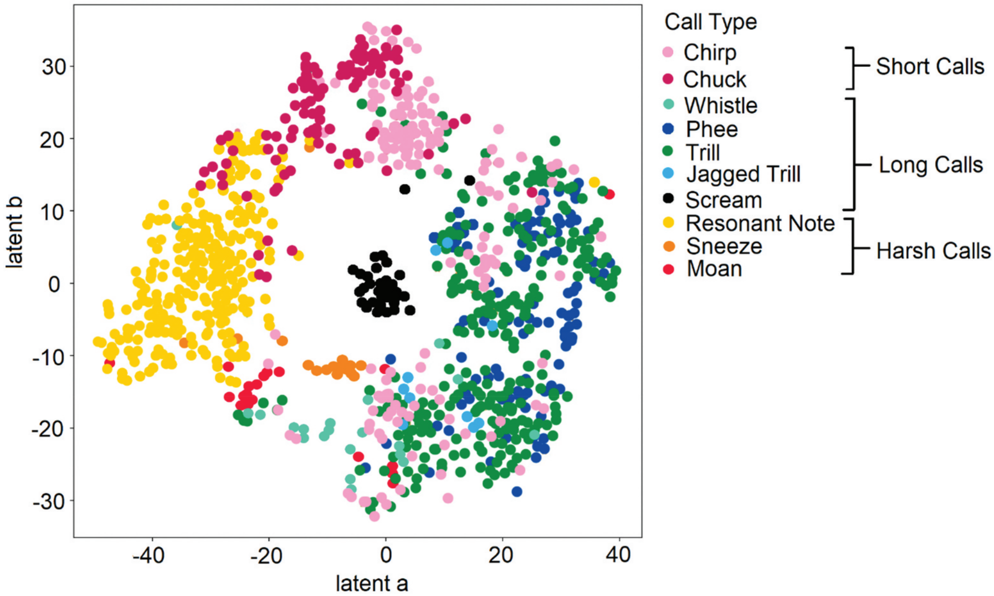

Figure 3. Results of the t-SNE clustering for all ten call types recorded from captive coppery titi monkeys, *Plecturocebus cupreus*, at Banham Zoo.

sneeze calls (Figure 2(h-j). All harsh calls were broadband and noisy in quality, showing no modulation. Sneezes are very short broadband calls and sound similar to a human sneeze. Moans are the lowest frequency call, often lost to background noise and graded with resonating notes. Resonating notes appeared as bisyllablic calls, with an inhale and exhale component. Sequences of these resonating notes formed the duet sequence. These calls showed deterministic chaos to varying degrees.

Many calls in this repertoire grade into each other. Chirps grade with chucks, trills and whistles, while chucks grade with resonating notes. Trills, whistles and phees all grade into each other (phees into whistles, phees into trill, etc.). Screams exist on a gradient with jagged trills and trills, with screams presenting as a more intense variation of jagged trills. Call groupings were confirmed by the t-SNE, showing distinct clustering of call types (Figure 3). The t-SNE also displays how call types may grade into each other (i.e. the overlap in the distribution of call types). Common sources of non-target sounds that are visually similar to titi monkey calls are noted in Figure 2(k-o).

Table 1. Precision, recall, accuracy, and F1 score metrics for call detection in captive coppery titi monkeys, *Plecturocebus cupreus*, across Banham, Shaldon, and Blackpool zoos. All measures are between 0 (low performance) and 1 (high performance).

Zoo	Precision	Recall	Accuracy	F1 Score
Banham	0.98	0.96	0.95	0.97
Shaldon	0.97	0.94	0.92	0.95
Blackpool	0.99	0.98	0.98	0.99
Overall	0.98	0.96	0.96	0.97

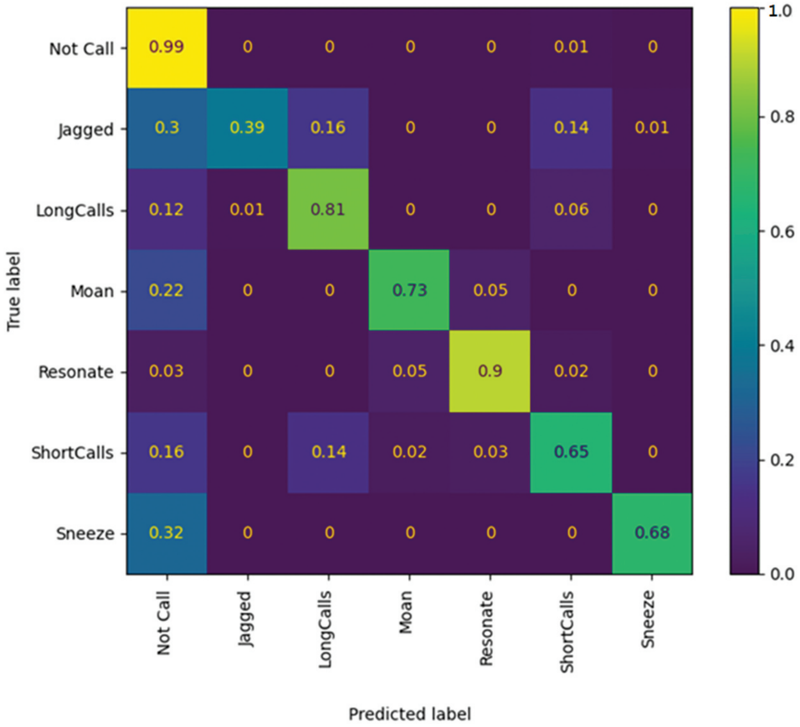

Figure 4. Confusion matrix showing overall classification accuracy across the seven combined call categories recorded in captive coppery titi monkeys, *Plecturocebus cupreus*. Not all categories were present at all zoos.

Detection

The model successfully detected *P. cupreus* calls across all three zoos (Table 1), with Blackpool Zoo data performing best across all metrics, while Shaldon Zoo had the lowest success across all metrics.

Table 2. Precision, recall, accuracy, and F1 score metrics for the seven condensed call categories recorded in captive coppery titi monkeys, *Plecturocebus cupreus*, across all zoos. All measures are between 0 (low performance) and 1 (high performance). The 'Short Calls' category includes Chirps and Chucks; the 'Long Calls' category includes Trills, Whistles and Phees, but not Jagged Trill (treated individually) or Scream (excluded due to small sample size); all 'Harsh Calls' are treated individually.

Call Category	Precision	Recall	Accuracy	F1 Score
No Call	0.96	0.97	0.95	0.97
Jagged Trill	0.65	0.39	0.98	0.48
Long Calls	0.82	0.80	0.96	0.81
Moan	0.58	0.73	0.99	0.65
Resonating Note	0.88	0.90	0.99	0.89
Short Calls	0.66	0.65	0.96	0.65
Sneeze	1.00	0.68	0.99	0.81

Classification

The model was able to accurately predict Resonating Notes, Moans, Long Calls and Non-Target Calls (absence of target call types), and was less successful with Sneezes, Short Calls and Jagged Trills (Figure 4, see also Supplementary Fig A for individual confusion matrices by zoo). For the overall sample, precision, recall, and F1 score was high for No Calls, Long Calls, Resonating Notes, and low for Moans and Jagged Trills (Table 2). Precision was 1.0 for Sneezes; however, this is likely due to the very small sample size of calls in the training data ($N = 23$). Accuracy was high for all call types (>0.95). Long Calls, Short Calls, Sneezes, Jagged Trills, and Moans were most frequently misclassified as no calls. Short Calls and Jagged Trills were also misclassified as Long Calls, and Short Calls and Long Calls were misclassified as each other.

Discussion

In this study we aimed to determine the vocal repertoire of *P. cupreus* and create an algorithm for detecting and classifying calls in streaming audio. The dataset used was unbalanced, small, and noisy, presenting a set of challenges common in machine learning (Arnaud et al. 2023). This was achieved by fitting two LSTM models, one for detection and one for classification. These were then evaluated using a series of metrics.

A vocal repertoire of ten calls was established, suggesting a smaller repertoire size than that described by Moynihan (1966), p. 12 calls) and Robinson (1979; 15 calls) for *P. moloch*. The repertoire was similar between species in terms of the call types observed, with both displaying chirps, chucks, sneezes, whistles, trills, screams, moans and resonating notes. Both species repertoires also displayed considerable gradation between calls, with all but sneeze calls grading with at least one other call type. Other studies of titi monkey calls have noted more diversity in call types, particularly in the alarm calls of *C. nigrifrons*, in which individuals used ‘A’ and ‘B calls’ in reference to different predators (Cäsar et al. 2013; Berthet et al. 2018). From a qualitative examination, chucks and chirps as used in our study of *P. cupreus* were most visually and descriptively similar to the ‘A calls’ of *C. nigrifrons*.

Our model was able to successfully detect and classify the simplified vocal repertoire of *P. cupreus* calls in streaming audio based on a small and unbalanced dataset with varying levels of noise. The detection element of the model performed similarly across all three zoo soundscapes but showed differing levels of effectiveness between zoos for the classification element. Differences between zoos may be a result of their specific soundscapes; the enclosures at Blackpool Zoo had little background noise due to soundproofed glass, while the enclosures at Banham Zoo and Shaldon Zoo were shared with other species and were not as sheltered from outside noise. Performance metrics were also variable across call types/categories, likely due to the substantial gradation between call types leading to call misclassifications, and from lower sample sizes for rarer calls such as jagged trills, moans, and sneezes. Lower precision scores for the prediction of Jagged Trills, Moans, and Short Calls likely indicates a tendency to false positives low precision in the model. This could be reduced by increasing the dataset size, for Jagged Trills and Moans. These results are similar in performance to other models developed to detect/classify common marmoset, *Callithrix jacchus*, calls (Phaniraj 2022), indri, *Indri indri*, song (Ravaglia et al. 2023), and grey mouse lemur, *Microcebus murinus*, calls (Romero-Mujalli et al. 2021).

Applications

General applications for the model could include more efficient processing of any acoustic data from *P. cupreus*, reducing time dedicated to detection and labelling. As *P. cupreus* showed similar call types to *P. moloch*, there is potential for testing the model on other titi species (*Callicebus* spp., *Cheracebus* spp., and *Plecturocebus* spp.), many of which are endangered and in need of population monitoring and further study (IUCN 2023). Acquiring titi monkey vocal data for training may be difficult due to decreasing wild populations and limited availability/absence in databases such as the Macaulay Library, and so being able to use captive *P. cupreus* calls would offer a potential alternative. However, when doing this, audio used between the two species must be similar in signal-to-noise ratio, sampling rate and normalisation in order to be effective (Kershenbaum et al. 2025).

Passive acoustic monitoring (PAM) involves the use of autonomous audio recorders to gather data non-invasively and has become increasingly vital to conservation operations in recent years (Browning et al. 2017; Hoefer et al. 2023). It is also a method that generates large amounts of data and so would benefit from being used alongside ML solutions. PAM can be used for determining the presence or absence of a species, and may be particularly useful for titi species, which are often found in dense vegetation and can be difficult to observe visually (Kalan et al. 2015; Bezerra et al. 2017; Cole et al. 2022; Souza-Alves et al. 2023). Estimations of population density from PAM could also be processed more efficiently with ML and could provide key information on how titi monkey species use their habitat and respond to changes. Habitat loss and disease outbreaks are ongoing and time-sensitive threats for titi monkeys (Berthet et al. 2021; Souva-Alves et al. 2023) and require rapid processing of data to respond efficiently with conservation actions.

ML applications within captive populations could include monitoring of repertoire use as a non-invasive means of assessing welfare. Call types in *P. moloch* have been associated with distress or hostility, including screams and chucks (Moynihan 1966; Robinson 1979). *C. nigrifrons* have also been found to alter their vocal behaviour, call less often and change their call frequencies in response to anthropogenic noise (Duarte et al. 2018). The model could be adapted to detect and classify calls, highlighting if certain call types occur over a given threshold. General advancement in the understanding of titi monkey behaviour, distribution, and ecology is first required for effective conservation and welfare actions (Souza-Alves et al. 2023).

Model improvements

The model could be improved by using data from wild populations, which would allow for a more complete repertoire and more authentic soundscape for the model to be trained on. Semi-supervised learning utilising the unlabelled data gathered (approx. 550 hours) could also make the model more robust by adding more training data (Reddy et al. 2018; Zhu and Goldberg 2022). For example, a form of self-learning /pseudo-labelling in which the current model's initial outputs are used as pseudo-data for a larger version of the model could be run (Yafen et al. 2022). Fine-tuning other existing models with our dataset, potentially using pre-trained feature extractors such as wav2vec2 (Baevski et al. 2020) may provide additional features other than MFCCs and boost model performance. Adapting one pretrained model (in this case, the current *P. cupreus* model) to fit another context (i.e. wild populations or different species) can, however, be challenging as the new audio data must closely resemble the original model's training data (i.e. sound-to-noise ratio, sampling rates; Kershenbaum et al. 2025).

Future studies

One aspect in which the model could be further developed is to identify individual callers. *Plecturocebus* spp. produce pair distinct duets (Muller and Anzenberger, 2002; Lau et al. 2020), which could be utilised in population monitoring or presence studies to distinguish family groups and their movements over time. Moynihan (1966) suggested that chirps may also be individually distinct; however, this remains to be tested. More detailed aspects of call structure could also be considered, such as the presence of nonlinear phenomena (NLP), which may relate to emotional arousal and has been previously connected to emotional valence (Briefer et al. 2015). In this case, being able to detect calls with NLP would give more insight into captive welfare. Expanding the vocal repertoire into wild populations would also allow for the model to be more effectively applied within population monitoring and presence/absence studies.

Conclusions

This study determined the vocal repertoire for *P. cupreus* and developed an effective initial ML model that was designed to detect, segment, and classify calls in streaming audio. This used a dataset typical of bioacoustic research, with a noisy and unbalanced sample with interference from other species. Our LSTM based model performed effectively across all metrics in all three zoos, detecting a condensed set of call categories (7 from the full repertoire of 10). This model could be applied within captive and wild settings, for monitoring captive and wild populations non-invasively. This would make data gathering and processing more feasible and efficient, which is essential with a species that is difficult to study in the wild and prone to illness in captive settings.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

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

References

Akbal E, Dogan S, Tuncer T. 2022. An automated multispecies bioacoustics sound classification method based on a nonlinear pattern: twine-pat. Ecol Inf. 68:101529.

Aldrich BC, Molleson L, Nekaris KAI. 2008. Vocalizations as a conservation tool: an auditory survey of the Andean titi monkey Callicebus oenanthe Thomas, 1924 (Mammalia: primates: pitheciidae) at Tarangue, Northern Peru. Contrib Zool. 77(1):1–6.

Alipek S, Maelzer M, Paumen Y, Schauer-Weisshahn H, Moll J. 2023. An efficient neural network design incorporating autoencoders for the classification of bat echolocation sounds. Animals. 13(16):2560.

Arnaud V, Pellegrino F, Keenan S, St-Gelais X, Mathevon N, Levréro F, Coupé C. 2023. Improving the workflow to crack Small, Unbalanced, Noisy, but Genuine (SUNG) datasets in bioacoustics: the case of bonobo calls. PLOS Comput Biol. 19(4):e1010325.

Baevski A, Zhou Y, Mohamed A, Auli M. 2020. wav2vec 2.0: a framework for self-supervised learning of speech representations. Adv Neural Inf Process Syst. 33:12449–12460.

Batist CH, Razafindraibe MN, Randriamanantena F, Baden AL. 2023. Bioacoustic characterization of the black-and-white ruffed lemur (*Varecia variegata*) vocal repertoire. Primates. 64(6):621–635.

Bergler C, Smeele SQ, Tyndel SA, Barnhill A, Ortiz ST, Kalan AK, Klump BC. 2022. ANIMAL-SPOT enables animal-independent signal detection and classification using deep learning. Sci Rep. 12(1):21966.

Bermant PC. 2021. BioCPPNet: automatic bioacoustic source separation with deep neural networks. Sci Rep. 11(1):23502.

Berthet M, Mesbahi G, Duvot G, Zuberbühler K, Cäsar C, Bicca-Marques JC. 2021. Dramatic decline in a titi monkey population after the 2016–2018 sylvatic yellow fever outbreak in Brazil. Am J Primatol. 83(12):e23335.

Berthet M, Neumann C, Mesbahi G, Cäsar C, Zuberbühler K. 2018. Contextual encoding in titi monkey alarm call sequences. Behavioral Ecol Sociobiol. 72:1–11.

Bezerra B, Casar C, Jerusalinsky L, Barnett A, Bastos M, Souto A, Jones G. 2017. Pitheciid vocal communication: what can we say about what they are saying? Ethnobiol Conserv. 6. https://ethnobioconservation.com/index.php/ebc/article/view/130.

Bianco MJ, Gerstoft P, Traer J, Ozanich E, Roch MA, Gannot S, Deledalle CA. 2019. Machine learning in acoustics: theory and applications. J Acoustical Soc Am. 146(5):3590–3628.

Boersma P, Weenink D. 2024. Praat: doing phonetics by computer [Computer program]. Version 6.4.27.

Briefer EF, Maigrot AL, Mandel R, Freymond SB, Bachmann I, Hillmann E. 2015. Segregation of information about emotional arousal and valence in horse whinnies. Sci Rep. 5(1):9989.

Briefer EF, Sypherd CCR, Linhart P, Leliveld LM, Padilla de la Torre M, Read ER, Tallet C. 2022. Classification of pig calls produced from birth to slaughter according to their emotional valence and context of production. Sci Rep. 12(1):3409.

Browning E, Gibb R, Glover-Kapfer P, Jones KE. 2017. *Passive acoustic monitoring in ecology and conservation*. WWF-UK. Doi:10.25607/OBP-876.

Cakir E, Adavanne S, Parascandolo G, Drossos K, Virtanen T. 2017 Aug. Convolutional recurrent neural networks for bird audio detection. In: 2017 25th European Signal Processing Conference (EUSIPCO); Kos, Greece. IEEE. p. 1744–1748.

Carroll BT, Anderson DV, Daley W, Harbert S, Britton DF, Jackwood MW. 2014 Dec. Detecting symptoms of diseases in poultry through audio signal processing. In: 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP); IEEE. p. 1132–1135.

Cäsar C. 2012. *Anti-predator behaviour of black-fronted titi monkeys (Callicebus nigrifrons)* [(doctoral dissertation]. University of St Andrews.

Cäsar C, Byrne R, Young RJ, Zuberbühler K. 2012. The alarm call system of wild black-fronted titi monkeys, *Callicebus nigrifrons*. Behav Ecol Sociobiol. 66:653–667.

Cäsar C, Zuberbühler K, Young RJ, Byrne RW. 2013. Titi monkey call sequences vary with predator location and type. Biol Lett. 9(5):20130535.

Chao KW, Hu NZ, Chao YC, Su CK, Chiu WH. 2019. Implementation of artificial intelligence for classification of frogs in bioacoustics. Symmetry. 11(12):1454.

Chung Y, Lee J, Oh S, Park D, Chang HH, Kim S. 2013. Automatic detection of cow's oestrus in audio surveillance system. Asian-Australas J Anim Sci. 26(7):1030.

Clink DJ, Crofoot MC, Marshall AJ. 2019. Application of a semi-automated vocal fingerprinting approach to monitor Bornean gibbon females in an experimentally fragmented landscape in Sabah, Malaysia. Bioacoustics. 28(3):193–209.

Coffey KR, Marx RE, Neumaier JF. 2019. DeepSqueak: a deep learning-based system for detection and analysis of ultrasonic vocalizations. Neuropsychopharmacology. 44(5):859–868.

Cole JS, Michel NL, Emerson SA, Siegel RB. 2022. Automated bird sound classifications of long-duration recordings produce occupancy model outputs similar to manually annotated data. Ornithol Appl. 124(2):duac003.

Diep D, Nonon H, Marc I, Delhom J, Roure F. 2013 Jun. Acoustic counting and monitoring of shad fish populations. International AmiBio Workshop: Recent Progress in Computational Bioacoustics for Assessing Biodiversity; Bonn, Germany.

Duan G, Zhang S, Lu M, Okinda C, Shen M, Norton T. 2021. Short-term feeding behaviour sound classification method for sheep using LSTM networks. Int J Agric Biol Eng. 14(2):43–54.

Duarte MHL, Kaizer MC, Young RJ, Rodrigues M, Sousa-Lima RS. 2018. Mining noise affects loud call structures and emission patterns of wild black-fronted titi monkeys. Primates. 59:89–97.

Garcia HA, Couture T, Galor A, Topple JM, Huang W, Tiwari D, Ratilal P. 2020. Comparing performances of five distinct automatic classifiers for fin whale vocalizations in beamformed spectrograms of coherent hydrophone array. Remote Sens. 12(2):326.

Gers FA, Schmidhuber J, Cummins F. 2000. Learning to forget: continual prediction with LSTM. Neural Comput. 12(10):2451–2471.

Goffinet J, Brudner S, Mooney R, Pearson J. 2021. Low-dimensional learned feature spaces quantify individual and group differences in vocal repertoires. Elife. 10:e67855.

Gong CSA, Su CHS, Chao KW, Chao YC, Su CK, Chiu WH. 2021. Exploiting deep neural network and long short-term memory method-ologies in bioacoustic classification of LPC-based features. PLOS ONE. 16(12):e0259140.

Grari M, Boukabous M, Yandouzi M, Berrahal M, Idrissi I, Moussaoui O, Moussaoui M. 2023. Forest fire detection and monitoring through environment sound spectrum using deep learning. J Theor Appl Inf Technol. 101(19):6080–6089.

Handelman GS, Kok HK, Chandra RV, Razavi AH, Huang S, Brooks M, Asadi H. 2019. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. Am J Roentgenol. 212(1):38–43.

Hill AP, Prince P, Piña Covarrubias E, Doncaster CP, Snaddon JL, Rogers A. 2018. AudioMoth: evaluation of a smart open acoustic device for monitoring biodiversity and the environment. Methods Ecol Evol. 9(5):1199–1211.

Hochreiter S, Schmidhuber J. 1997. Long short-term memory. Neural Comput. 9(8):1735–1780.

Hoefer S, McKnight DT, Allen-Ankins S, Nordberg EJ, Schwarzkopf L. 2023. Passive acoustic monitoring in terrestrial vertebrates: a review. Bioacoustics. 32(5):1–26.

Hossain S, Calloway C, Lippa D, Niederhut D, Shupe D. 2019 Jul. Visualization of bioinformatics data with Dash Bio. In: *18th Python in Science Conference*; Austin, Texas. p.126–133.

Huang HT, Downey AR, Bakos JD. 2022. Audio-based wildfire detection on embedded systems. Electronics. 11(9):1417.

Ibrahim AK, Zhuang H, Chérubin LM, Schärer Umpierre MT, Ali AM, Nemeth RS, Erdol N. 2019. Classification of red Hind grouper call types using random ensemble of stacked autoencoders. J Acoust Soc Am. 146(4):2155–2162.

Islam S, Valles D. 2020 Jan. Houston toad and other chorusing amphibian species call detection using deep learning architectures. 2020 10th Annual Computing and Communication Workshop and Conference (CCWC); IEEE. p. 511–516.

IUCN. 2023. The IUCN red list of threatened species. Version 2023-1. [accessed 2023 September 30]. https://www.iucnredlist.org .

IUCN. 2024. The IUCN red list of threatened species. Version 2022-2. https://www.iucnredlist.org.

Jiang Y, Liu Z, Yang C, White P, Wang X, Lü LG, Kittiwattanawong K. 2024. Dugong chirp type classification based on fundamental contour extraction and hierarchical cluster analysis. Appl Acoust. 217:109812.

Jo T. 2021. Machine learning foundations: supervised, unsupervised, and advanced learning. Springer International Publishing. doi:10.1007/978-3-030-65900-4.

Jones ML. 1968. Longevity of primates in captivity. Int Zoo Yearb. 8(1):183–192.

Jung DH, Kim NY, Moon SH, Jhin C, Kim HJ, Yang JS, Park SH. 2021. Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals. 11(2):357.

Kalan AK, Mundry R, Wagner OJ, Heinicke S, Boesch C, Kühl HS. 2015. Towards the automated detection and occupancy estimation of primates using passive acoustic monitoring. Ecol Indic. 54:217–226.

Katsis LK, Hill AP, Pina-Covarrubias E, Prince P, Rogers A, Doncaster CP, Snaddon JL. 2022. Automated detection of gunshots in tropical forests using convolutional neural networks. Ecol Indic. 141:109128.

Kershenbaum A, Akçay C, Babu-Saheer L, Barnhill A, Best P, Cauzinille J, Dunn JC. 2025. Automatic detection for bioacoustic research: a practical guide to the state of the art and future directions. Biol Rev. 100(2): 620–646.

Kinzey WG, Rosenberger AL, Heisler PS, Prowse DL, Trilling JS. 1977. A preliminary field investigation of the yellow handed titi monkey, *Callicebus torquatus torquatus*, in northern Peru. Primates. 18:159–181.

Koops HV, Van Balen J, Wiering F. 2015 Sept 8-11. Automatic segmentation and deep learning of bird sounds. Experimental IR Meets Multilinguality, Multimodality, and Interaction: 6th International Conference of the CLEF Association, CLEF'15, Toulouse, France; September 8-11; Springer International Publishing 6; p. 261–267.

Landis JR, Koch GG. 1977. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. Biometrics. 33:363–374.

Larranaga A, Bielza C, Pongrácz P, Faragó T, Bálint A, Larranaga P. 2015. Comparing supervised learning methods for classifying sex, age, context and individual Mudi dogs from barking Anim Cognition. 18(2):405–421.

Lau AR, Clink DJ, Bales KL. 2020. Individuality in the vocalizations of infant and adult coppery titi monkeys (*Plecturocebus cupreus*). Am J Primatol. 82(6):e23134.

Lever J. 2016. Classification evaluation: it is important to understand both what a classification metric expresses and what it hides. Nat Methods. 13(8):603–605.

Lorenz R, Mason WA. 1971. Establishment of a colony of titi monkeys. Int Zoo Yearb. 11 (1):168–174.

Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, Jones KE. 2018. Bat detective—deep learning tools for bat acoustic signal detection. PLOS Comput Biol. 14(3): e1005995.

Mao A, Giraudet CS, Liu K, De Almeida Nolasco I, Xie Z, Xie Z, McElligott AG. 2022. Automated identification of chicken distress vocalizations using deep learning models. J R Soc Interface. 19 (191):20210921.

Mao A, Mohri M, Zhong Y. 2023. Cross-entropy loss functions: theoretical analysis and applications. In: *International conference on Machine learning*; Hawaii, USA. PMLR. p. 23803–23828.

Martin K, Cornero FM, Clayton NS, Adam O, Obin N, Dufour V. 2024. Vocal complexity in a socially complex corvid: gradation, diversity and lack of common call repertoire in male rooks. R Soc Open Sci. 11(1):231713.

McFee B, Raffel C, Liang D, Ellis DP, McVicar M, Battenberg E, Nieto O. 2015 July. Librosa: audio and music signal analysis in python. In: *Proceedings of the 14th python in science conference* 8; p. 18–25.

McInnes L, Healy J, Melville J. 2018. UMAP: uniform manifold approximation and projection for dimension reduction. arXiv Prepr arXiv 1802 03426. https://arxiv.org/abs/1802.03426.

Mcloughlin MP, Stewart R, McElligott AG. 2019. Automated bioacoustics: methods in ecology and conservation and their potential for animal welfare monitoring. J R Soc Interface. 16 (155):20190225.

Mendoza SP, Mason WA. 1997. Attachment relationships in new world primates. Ann of the New Y Acad Sci. 807:203–209.

Merchan F, Guerra A, Poveda H, Guzmán HM, Sanchez-Galan JE. 2020. Bioacoustic classification of Antillean manatee vocalization spectrograms using deep convolutional neural networks. Appl Sci. 10(9):3286.

Morales G, Vargas V, Espejo D, Poblete V, Tomasevic JA, Otondo F, Navedo JG. 2022. Method for passive acoustic monitoring of bird communities using UMAP and a deep neural network. Ecol Inf. 72:101909.

Moynihan M. 1966. Communication in the titi monkey, *Callicebus*. J Zool. 150(1):77–127.

Müller AE, Anzenberger G. 2002. Duetting in the titi monkey *Callicebus cupreus*: structure, pair specificity and development of duets. Folia Primatol. 73(2–3):104–115.

Narasimhan R, Fern XZ, Raich R. 2017 Mar. Simultaneous segmentation and classification of bird song using CNN. *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE. p. 146–150.

Nasirahmadi A, Gonzalez J, Sturm B, Hensel O, Knierim U. 2020. Pecking activity detection in group-housed turkeys using acoustic data and a deep learning technique. Biosystems Eng. 194:40–48.

Nieto-Mora DA, Rodríguez-Buritica S, Rodríguez-Marín P, Martínez-Vargaz JD, Isaza-Narváez C. 2023. Systematic review of machine learning methods applied to ecoacoustics and soundscape monitoring. Heliyon. 9(10):e20275–e20275. doi: 10.1016/j.heliyon.2023.e20275.

Noda JJ, Travieso CM, Sánchez-Rodríguez D, Dutta MK, Singh A. 2016 Feb. Using bioacoustic signals and support vector machine for automatic classification of insects. *2016 3rd international conference on signal processing and integrated networks (SPIN)*; IEEE. p. 656–659.

Noumida A, Rajan R. 2022. Multi-label bird species classification from audio recordings using attention framework. Appl Acoust. 197:108901.

Ogundile OO, Babalola OP, Odeyemi SG, Rufai KI. 2022. Hidden Markov models for detection of mysticetes vocalisations based on principal component analysis. Bioacoustics. 31(6):710–738.

Oikarinen T, Srinivasan K, Meisner O, Hyman JB, Parmar S, Fanucci-Kiss A, Feng G. 2019. Deep convolutional network for animal sound classification and source attribution using dual audio recordings. J Acoustical Soc Am. 145(2):654–662.

Oswald JN, Erbe C, Gannon WL, Madhusudhana S, Thomas JA. 2022. Detection and classification methods for animal sounds. Exploring Anim Behav Through Sound. 1:269–317.

Ozanich E, Thode A, Gerstoft P, Freeman LA, Freeman S. 2021. Deep embedded clustering of coral reef bioacoustics. J Acoust Soc Am. 149(4):2587–2601.

Parada P, Cardenal-López A. 2014. Using Gaussian mixture models to detect and classify dolphin whistles and pulses. J Acoustical Soc Am. 135(6):3371–3380.

Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Chintala S. 2019. Pytorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst. 32:8024–8035.

Paumen Y, Mälzer M, Alipek S, Moll J, Lüdtke B, Schauer-Weisshahn H. 2022. Development and test of a bat calls detection and classification method based on convolutional neural networks. Bioacoustics. 31(5):505–516.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Duchesnay É. 2011. Scikit-learn: machine learning in python. J Mach Learn Res. 12:2825–2830.

Phaniraj N. 2022. Machine learning approaches for tracking changes in marmoset vocalizations and social behaviors [doctoral dissertation]. Indian Institute of Science Education and Research.

Pinto, L. P., Barnett, A. A., Bezerra, B. M., Boubli, J. P., Bowler, M., Cardoso, N. D. A. & Veiga, L. M. (2013). Why we know so little: the challenges of fieldwork on the Pitheciids. In: *Evolutionary biology and conservation of titis, sakis and uacaris. Cambridge: Cambridge University Press*; p. 145.

Ravaglia D, Ferrario V, De Gregorio C, Carugati F, Raimondi T, Cristiano W, Gamba M. 2023. There you are! Automated detection of Indris' songs on features extracted from passive acoustic recordings. Animals. 13(2):241.

Reddy YCAP, Viswanath P, Reddy BE. 2018. Semi-supervised learning: a brief review. Int J Eng Technol. 7(1.8):81.

Robinson JG. 1979. An analysis of the organization of vocal communication in the titi monkey *Callicebus moloch*. Z für Tierpsychol. 49(4):381–405.

Romero-Mujalli D, Bergmann T, Zimmermann A, Scheumann M. 2021. Utilizing DeepSqueak for automatic detection and classification of mammalian vocalizations: a case study on primate vocalizations. Sci Rep. 11(1):24463.

Schön PC, Puppe B, Manteuffel G. 2001. Linear prediction coding analysis and self-organizing feature map as tools to classify stress calls of domestic pigs (*Sus scrofa*). J Acoustical Soc Am. 110 (3):1425–1431.

Schröter H, Nöth E, Maier A, Cheng R, Barth V, Bergler C. 2019 May. Segmentation, classification, and visualization of orca calls using deep learning. ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) IEEE. p. 8231–8235.

Shen W, Ji N, Yin Y, Dai B, Tu D, Sun B, Zhao Y. 2022. Fusion of acoustic and deep features for pig cough sound recognition. Comput Electron Agriculture. 197:106994.

Somwong B, Kumphet K, Massagram W. 2023 June. Acoustic monitoring system with AI threat detection system for forest protection. *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*; IEEE. p. 253–257.

Souza-Alves JP, Boyle SA, Barnett AA. 2023. Knowledge shortfalls for titi monkey: a poorly known clade of small-bodied South American primates. Biol Conserv. 286:110256.

Srisuphab A, Kaakkurivaara N, Silapachote P, Tangkit K, Meunpong P, Sunetnanta T. 2020 Nov. Illegal logging listeners using IoT networks. *2020 IEEE REGION 10 CONFERENCE (TENCON)*; IEEE. p. 1277–1282.

Srujana C, Sriya B, Divya S, Shaik S, Kakulapati V. 2023 June. Species identification of birds via acoustic processing signals using recurrent network analysis (RNN). In *International Conference on Soft Computing and Signal Processing*; Singapore: Springer Nature Singapore; p. 27–38.

Steinfath E, Palacios-Muñoz A, Rottschäfer JR, 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.

Sugali K, Sprunger C, Inukollu VN. 2024. AI testing: ensuring a good data split between data sets (training and test) using K-means clustering and decision tree analysis. Int J Surg Pathol On Soft Computing. 15(1):34–45.

Tallet C, Linhart P, Policht R, Hammerschmidt K, Šimeček P, Kratinova P, Špinka M. 2013. Encoding of situations in the vocal repertoire of piglets (*Sus scrofa*): a comparison of discrete and graded classifications. PLOS ONE. 8(8):e71841.

Thomas M, Jensen FH, Averly B, Demartsev V, Manser MB, Sainburg T, Strandburg-Peshkin A. 2022. A practical guide for generating unsupervised, spectrogram-based latent space representations of animal vocalizations. The J Anim Ecol. 91(8):1567–1581.

Townsend JP, Aldstadt J. 2023. Habitat suitability mapping using logistic regression analysis of long-term bioacoustic bat survey dataset in the Cassadaga Creek watershed (USA). Sci Total Environ. 895:165077.

Trapanotto M, Nanni L, Brahnam S, Guo X. 2022. Convolutional neural networks for the identification of African lions from individual vocalizations. J Imag. 8(4):96.

Trawicki MB. 2021. Multispecies discrimination of whales (cetaceans) using hidden Markov models (HMMs). Ecol Inf. 61:101223.

Tseng YC, Eskelson BN, Martin K, LeMay V. 2021. Automatic bird sound detection: logistic regression based acoustic occupancy model. Bioacoustics. 30(3):324–340.

Valente D, De Gregorio C, Torti V, Miaretsoa L, Friard O, Randrianarison RM, Gamba M. 2019. Finding meanings in low dimensional structures: stochastic neighbor embedding applied to the analysis of Indri indri vocal repertoire. Animals. 9(5):243.

Vamsi B, Mahanty M, Doppala BP. 2023. An auto encoder-decoder approach to classify the bird sounds using deep learning techniques. SN Comput Sci. 4(3):289.

Van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. J Mach Learn Res. 9 (11):2579–2605.

Van Kuijk SM, O'Brien S, Clink DJ, Blake JG, 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. Front Ecol Evol. 11:11. doi: 10.3389/fevo.2023.1173722.

Veiga LM, Barnett AA, Ferrari SF, Norconk MA. 2013. Evolutionary biology and conservation of titis, sakis and uacaris Vol. 65. Cambridge: Cambridge University Press.

Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, Van Mulbregt P. 2020. SciPy 1.0: fundamental algorithms for scientific computing in python. Nat Methods. 17 (3):261–272.

Wadewitz P, Hammerschmidt K, Battaglia D, Witt A, Wolf F, Fischer J. 2015. Characterizing vocal repertoires—hard vs. soft classification approaches. PLOS ONE. 10(4):e0125785.

Wang H, Xu Y, Yu Y, Lin Y, Ran J. 2022. An efficient model for a vast number of bird species identification based on acoustic features. Animals. 12(18):2434.

Xie J, Towsey M, Zhang J, Roe P. 2020. Investigation of acoustic and visual features for frog call classification. J Signal Process Syst. 92:23–36.

Xie J, Zhu M, Hu K, Zhang J, Hines H, Guo Y. 2022. Frog calling activity detection using lightweight CNN with multi-view spectrogram: a case study on Kroombit tinker frog. Mach Learn Appl. 7:100202.

Xu W, Zhang X, Yao L, Xue W, Wei B. 2020. A multi-view CNN-based acoustic classification system for automatic animal species identification. Ad Hoc Networks. 102:102115.

Yafen LIU, Yifeng ZH, Lingyi JI, Guohe LI, Wenjie ZH. 2022. Survey on pseudo-labeling methods in deep semi-supervised learning. J Front Comput Sci Technol. 16(6):1279.

Zhang Z, Sabuncu M. 2018. Generalized cross entropy loss for training deep neural networks with noisy labels. Adv Neural Inf Process Syst. 31:8792–8802.

Zhao Z, Zhang SH, Xu ZY, Bellisario K, Dai NH, Omrani H, Pijanowski BC. 2017. Automated bird acoustic event detection and robust species classification. Ecol Inf. 39:99–108.

Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, Zhang H. 2021. Machine learning: new ideas and tools in environmental science and engineering. Environ Sci; Technol. 55(19):12741–12754.

Zhu X, Goldberg AB. 2022. Introduction to semi-supervised learning. Springer Nature.

Znidersic E, Towsey M, Roy WK, Darling SE, Truskinger A, Roe P, Watson DM. 2020. Using visualization and machine learning methods to monitor low detectability species—the least bittern as a case study. Ecol Inf. 55:101014.

Zsebők S, Herczeg G, Laczi M, Nagy G, Vaskuti É, Hargitai R, Garamszegi LZ. 2021. Sequential organization of birdsong: relationships with individual quality and fitness. Behav Ecol. 32 (1):82–93.

Zualkernan I, Judas J, Mahbub T, Bhagwagar A, Chand P. 2021 Jan. An AIOT system for bat species classification. *2020 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS)*; IEEE. p. 155–160.