Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Original Articles

Automated detection of gunshots in tropical forests using convolutional neural networks

Lydia K.D. Katsis[a,*], Andrew P. Hill[b], Evelyn Piña-Covarrubias[a], Peter Prince[b], Alex Rogers[c], C. Patrick Doncaster[d], Jake L. Snaddon[a]

[a] *School of Geography and Environmental Science, Faculty of Environmental and Life Sciences, University of Southampton, Southampton, UK*
[b] *Agents, Interactions and Complexity, Electronics and Computer Science, Faculty of Physical Sciences and Engineering, University of Southampton, Southampton, UK*
[c] *Department of Computer Science, University of Oxford, Oxford, UK*
[d] *School of Biological Sciences, Institute for Life Sciences, University of Southampton, Southampton, UK*

ARTICLE INFO

Keywords:
Acoustic monitoring
Gunshots
Machine learning
Wildlife conservation
Tropical forests

ABSTRACT

Unsustainable hunting is one of the leading drivers of global biodiversity loss, yet very few direct measures exist due to the difficulty in monitoring this cryptic activity. Where guns are commonly used for hunting, such as in the tropical forests of the Americas and Africa, acoustic detection can potentially provide a solution to this monitoring challenge. The emergence of low cost autonomous recording units (ARUs) brings into reach the ability to monitor hunting pressure over wide spatial and temporal scales. However, ARUs produce immense amounts of data, and long term and large-scale monitoring is not possible without efficient automated sound classification techniques. We tested the effectiveness of a sequential two-stage detection pipeline for detecting gunshots from acoustic data collected in the tropical forests of Belize. The pipeline involved an on-board detection algorithm which was developed and tested in a prior study, followed by a spectrogram based convolutional neural network (CNN), which was developed in this manuscript. As gunshots are rare events, we focussed on developing a classification pipeline that maximises recall at the cost of increased false positives, with the aim of using the classifier to assist human annotation of files. We trained the CNN on annotated data collected across two study sites in Belize, comprising 597 gunshots and 28,195 background sounds. Predictions from the annotated validation dataset comprising 150 gunshots and 7044 background sounds collected from the same sites yielded a recall of 0.95 and precision of 0.85. The combined recall of the two-step pipeline was estimated at 0.80. We subsequently applied the CNN to an un-annotated dataset of over 160,000 files collected in a spatially distinct study site to test for generalisability and precision under a more realistic monitoring scenario. Our model was able to generalise to this dataset, and classified gunshots with 0.57 precision and estimated 80% recall, producing a substantially more manageable dataset for human verification. Using a classifier-guided listening approach such as ours can make wide scale monitoring of threats such as hunting a feasible option for conservation management.

1. Introduction

Biodiversity is being lost globally at an unprecedented rate, due to accelerating human impacts (Barnosky et al., 2011; Ceballos et al., 2020; IPBES, 2019). Species overexploitation is amongst the leading drivers of global biodiversity loss, threatening more than a quarter of all terrestrial animal species (WWF, 2016). This threat is expressed particularly forcefully in tropical regions, and it frequently occurs within protected areas (Jones et al., 2018; Laurance et al., 2012), where threats such as habitat loss commonly co-occur with overhunting to exacerbate the issue (Peres, 2001).

In contrast to the other leading drivers of global biodiversity loss such as habitat degradation (Hansen et al., 2013), there is limited spatiotemporal information on hunting. Traditional remote sensing methods have provided a wealth of high resolution spatial data on large-scale forest loss; these methods, however, cannot distinguish 'empty' forests that are structurally intact but devoid of fauna from truly healthy ecosystems (Benítez-López et al., 2019; Peres et al., 2006). Attempts to

[*] Corresponding author. Present address: Building 44 University of Southampton University Road, Southampton, SO17 1BJ, UK.
E-mail address: l.k.d.katsis@soton.ac.uk (L.K.D. Katsis).

https://doi.org/10.1016/j.ecolind.2022.109128
Received 17 December 2021; Received in revised form 26 February 2022; Accepted 28 June 2022
Available online 4 July 2022
1470-160X/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

map hunting over large spatial scales have consequently relied on predictors such as landscape accessibility (Benítez-López et al., 2019; Ziegler et al., 2016), yet these relationships are rarely validated on the ground (Deith and Brodie, 2020). Spatially explicit measures of hunting on a finer scale have commonly been obtained from ranger patrol data collected using law-enforcement monitoring software (Critchlow et al., 2017; Critchlow et al., 2015; Hötte et al., 2016; Plumptre et al., 2014) and systematic camera trapping grids (Ferreguetti et al., 2018; Hossain et al., 2016). Encounter data from patrols present challenges due to the deliberate bias of patrol effort towards areas with perceived high levels of activity (Dobson et al., 2020; Dobson et al., 2019) combined with avoidance of patrol routes by hunters, while camera traps have a limited field of detection in dense forests and are prone to theft.

Advances in the field of automated acoustic monitoring have opened up new avenues for directly monitoring anthropogenic disturbance (Astaras et al., 2017; Dobbins et al., 2020; Wrege et al., 2017) and biodiversity (Pijanowski et al., 2011; Sethi et al., 2020a; Sugai et al., 2018) in a systematic and spatially explicit manner. Acoustic monitoring captures a breadth of information from a variety of vocalising taxa (Bergler et al., 2019; Do Nascimento, 2020; Dufourq et al., 2021; Wrege et al., 2017), and from human activities such as logging and gun-hunting (Dobbins et al., 2020; Prince et al., 2019; Sethi et al., 2020b). The advent of low-cost autonomous recording units (ARUs: Hill et al., 2019) has enabled monitoring over large spatial and temporal scales, including in remote areas, providing a particularly effective method for monitoring cryptic species or activities (Campos-Cerqueira and Aide, 2016; Dobbins et al., 2020; Picciulin et al., 2019), while the overall soundscape can provide insight into habitat quality and biodiversity health (Sethi et al., 2020b).

Widescale uptake of this technology for ecological monitoring and conservation management is limited by users' capacity to analyse the information-dense data produced by the recorders. ARUs collect vast quantities of data in short time periods, which require processing to extract information on detections of acoustic events. The majority of ecological studies implementing acoustic monitoring currently rely on manual techniques such as listening to recordings and visually scanning spectrograms (Sugai et al., 2018), however these methods can become prohibitively time consuming with larger datasets. Consequently automated classification methods are a prerequisite for large-scale monitoring (Priyadarshani et al., 2018).

Current automated gunshot detection methods have several major shortcomings. Firstly, development of methods has largely relied on a carefully curated dataset of idealised gunshots with minimum background noise and high SNR (signal to noise ratio), which is rarely representative of field recordings from forested environments. Consequently, many of these algorithms can recognise specific examples but are unlikely to generalise to a realistic dataset (Nimmy et al., 2018; Valenzise et al., 2007). This issue is particularly problematic given the reliance on cross correlation and template matching schemes (Nimmy et al., 2018; Van der Merwe and Jordaan, 2013; Wrege et al., 2017), which are often highly sensitive to noise and fail to generalize (Nimmy et al., 2018).

Secondly, many of these approaches do not consider the extensive and variable nature of background sounds present in natural settings such as a tropical forests, or the relative rarity of target events such as gunshots compared to background sounds in a realistic monitoring scenario, which applies for both forested environments and urban environments (Chacon-Rodriguez et al., 2011; Hrabina et al., 2016; Singh et al., 2020). A wide range of background sounds can easily be confused with a gunshot, such as a woodpecker drumming, or a branch cracking, which could easily outnumber the much rarer true gunshot detections, resulting in a classifier with low precision (Wrege et al., 2017).

Thirdly, reliance on proprietary sound-analysis software and tools has resulted in a lack of transparency of methodology and unknown performance metrics. For example, Dobbins et al. 2020 used a combination of clustering and machine learning tools in the proprietary Kaleidoscope 5 Pro software (Wildlife Acoustics Inc.) to identify gunshots from field data collected in forests of Belize. However, there is no publicly available information on the specific methods used or the performance of this classifier, such as recall (proportion of gunshots detected) or precision (proportion of correctly predicted gunshots out of all gunshot predictions), which underpin the validity of the approach. Similarly, there is a lack of empirical evidence to validate other proprietary acoustic gunshot detection systems, such as the tool provided by Rainforest Connection (RFCx), and the 'Shotspotter' software which is widely used in urban areas across the US.

The recent development of spectrogram based convolutional neural networks (CNNs) has proved a powerful approach to automate detection of sound events (Dufourq et al., 2021; Kahl et al., 2021; Liu et al., 2019; Ruff et al., 2021), and provides a potential solution for robust gunshot detection (Bajzik et al., 2020; Khunarsa et al., 2010; Morehead et al., 2019; Mushtaq and Su, 2020). The general methodology involves converting the sounds to images of spectrograms, which represent the signal frequency and amplitude over time, and training an image-based CNN to classify the spectrogram images. This approach has proved very successful for complex animal vocalisations that are easily distinguished by humans, although it does not perform as well for more simple sounds that are easily confused with background noise (Bergler et al., 2019; Florentin et al., 2020), and consequently it has uncertain utility for detecting gunshot sounds in noisy tropical forests. Use of CNNs has been investigated for gunshot detection in urban areas (Bajzik et al., 2020; Khunarsa et al., 2010; Morehead et al., 2019; Mushtaq and Su, 2020), but not with a realistic field dataset form a tropical forest environment. While case studies based on urban gunshots report metrics of high precision and recall, they lack validation on real world data, which is critical to gauge their utility in real world monitoring scenarios (Bajzik et al., 2020; Morehead et al., 2019; Mushtaq and Su, 2020). Publicly available, annotated audio datasets collected in the field are essential for the development of improved classifiers, and to provide a benchmark for validation and comparative tests of existing algorithms. However there is a lack of publicly archived audio data, especially from tropical regions (Gibb et al., 2019).

In this study, we investigate the feasibility of using CNNs to detect the presence of acoustic gunshot events within tropical forests. As gunshots are rare events, we specifically aimed to create a classifier that maximises recall of gunshots (and therefore minimises false negatives) even at the cost of increased false positives, with the purpose of assisting human annotation of sound files as opposed to fully automated classification. We implemented a two-stage classification pipeline involving an on-board gunshot detector and CNN spectrogram image classification. The onboard detector was developed and tested in a previous study (Prince et al., 2019); this study develops and tests the post-hoc classification model. To train and validate our CNN we compiled an annotated training and validation dataset from data collected in two study sites in central Belize, with additional ground-truthed gunshots from a field test, which we have made openly accessible with this manuscript. We implemented our final model on a much larger dataset of raw, unannotated field recordings collected from a separate site in southern Belize to gauge the true precision of the classifier when tested on data with a more realistic distribution of negative files.

2. Methods

2.1. Study sites

Data were collected in three study sites across Belize's protected areas network (Fig. 1). Site 1, Tapir Mountain Nature Reserve (TMNR) is a small non-extractive nature reserve (IUCN category Ia), that consists of lowland broad-leaved moist forest on rugged karst hills. Adjacent to TMNR is Pook's Hill Reserve, where the field test of ground-truthed gunshots was conducted. Site 2 consists of a network of protected areas including Manatee Forest Reserve and several smaller

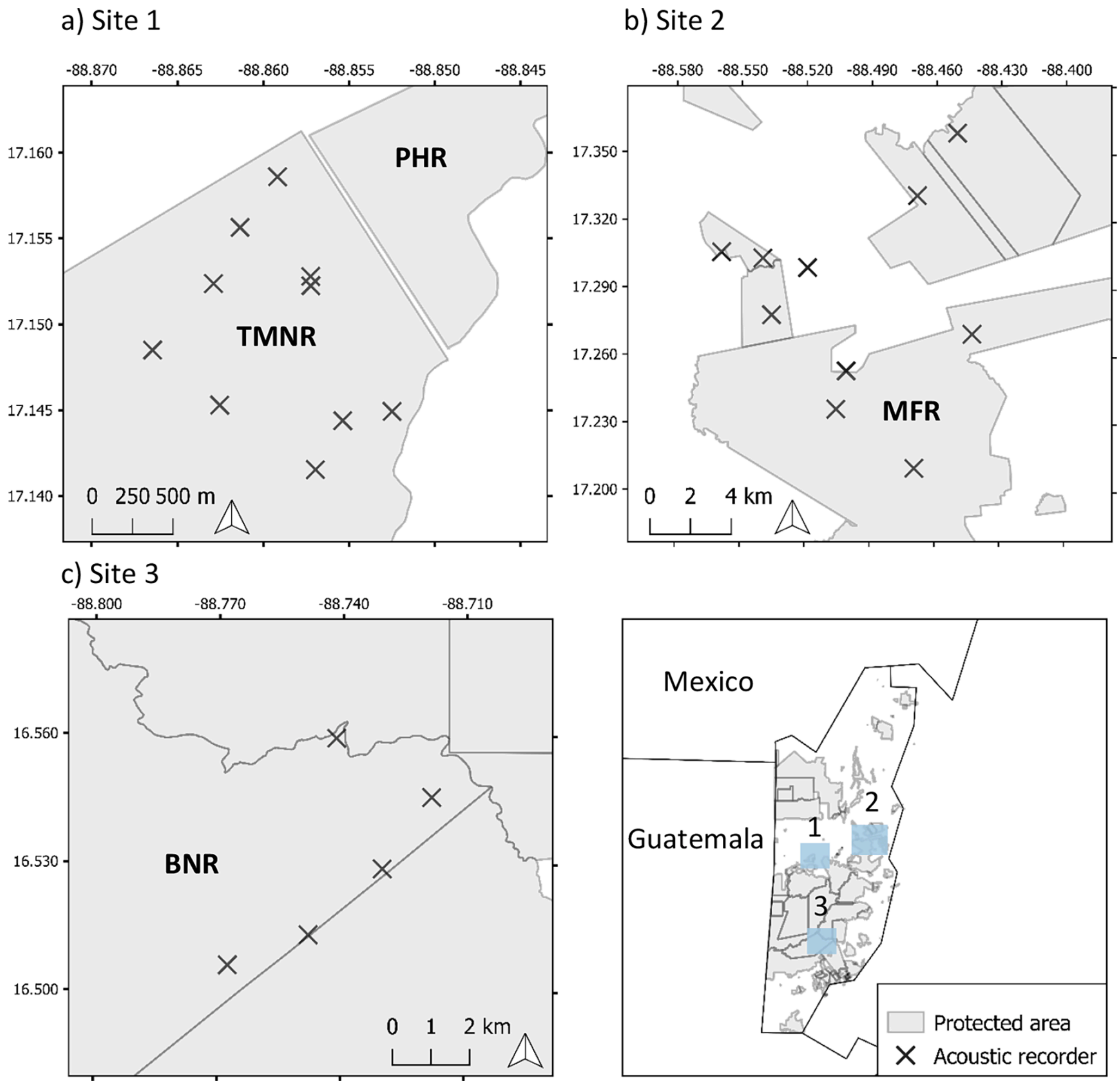

Fig. 1. Acoustic recorder deployment sites in Belize, indicated by blue boxes in the bottom right map. Site 1 is Tapir Mountain Nature Reserve (TMNR), which is adjacent Pook's Hill Reserve (PHR), Site 2 consists of Manatee Forest Reserve and surrounding smaller protected areas and private property, and Site 3 is Bladen Nature Reserve (BNR). Sites 1 and 2 contributed training data for the CNN, and the CNN was then implemented on data collected in Site 3. Note site maps are projected in WGS 84 and are of different scales.

neighbouring protected areas including Monkey Bay Wildlife Sanctuary, Peccary Hills National Park, Runaway Creek Nature Reserve, and several private properties. These reserves are assigned a range of less strict protection statuses than Site 1, although public access and hunting is prohibited in all of them. Primary ecosystems consist of lowland broadleaved moist forest, lowland savannah, and lowland pine forest on steep limestone hills. Site 3, Bladen Nature Reserve (BNR), is a strictly protected nature reserve (IUCN category Ia) in the south of the country. BNR encompasses some of the most pristine and biodiverse forests in the country, including broadleaf forests, savannah, and submontane forest, on limestone and granite hills.

Gun hunting is prevalent in all three study areas, despite formal protection. As is typical in the Neotropics, hunting largely targets medium and large bodied mammals, including peccaries, deer, paca, and armadillo (Foster et al., 2016). Gun hunting in Belize typically involves the use of shotguns as opposed to rifles, and subsequent analyses focus on the gunshot sound produced by 12-gauge and 16-gauge shotguns of the type typically used in Belize.

2.2. Data collection

Acoustic monitoring was conducted in Site 1 between March 2018 – March 2019 using AudioMoth acoustic sensors (v 1.0.0), modified to connect to a 6 V alkaline lantern battery, and housed in a rubber-sealed watertight plumbing tube with a Schlegel acoustic vent (Hill et al., 2019). AudioMoths were deployed at ten locations determined by a greedy heuristic algorithm, which predicted optimal placements for maximising the probability of detecting gunshots anywhere in the reserve (Fig. 1; see Pina-Covarrubias et al. 2019 for further information on study design).

Acoustic monitoring was conducted in Sites 2 and 3 from January 2021 – June 2021. This survey period was chosen as it coincided with an existing camera trapping survey. In these sites we used AudioMoth acoustic sensors (v 1.0.0 and v 1.1.0) with lithium AA batteries and housed in the AudioMoth IPX7 waterproof case, with a Schlegel acoustic vent. We chose to use data from just 1 month at each site, from ten locations in Site 1, and five locations in Site 3 (Fig. 1). AudioMoth deployment in these sites did not follow the optimisation process used in Site 1. Sensors in these sites were instead deployed alongside a camera trapping project, which followed a 2 km grid using existing logging roads, trails, and newly cut trails.

At each location, we deployed a pair of AudioMoths, one configured to record during the day (07:00 – 17:00 local time), and one during the night (17:00 – 07:00). We attached AudioMoths to the nearest suitable tree at shoulder height and directed the sensor towards the valley. AudioMoths were flashed with a custom firmware developed to detect and record gunshots in Belizean tropical forests (see Prince et al. 2019). Predicted gunshot-like sounds are recorded as 4.09 s Waveform Audio File Format (wav) files, collected at a sample rate of 8 kHz. To reduce the number of false positives, a threshold of 100 recordings per hour was set, after which the device went to sleep for the rest of the hour. Recall of this detector was 0.84 for gunshots fired within 500 m of the sensor; beyond this distance factors such as terrain and foliage caused a significant decrease in recall (Prince et al., 2019). The classification pipeline presented is thus applied within the context of gunshots occurring up to 500 m from the sensor.

We boosted the training dataset with a set of recordings of controlled gunshots collected in 2017 in Pook's Hill Reserve. This dataset contained gunshots fired during the day and the night at distances of 0–1000 m from 13 sensors. Continuous recordings were divided into 4 s sections and manually classified as containing gunshot or background sound, producing a dataset of 357 gunshot recordings. Although the time of the gunshot was known in the ground-truthing exercise, there was still a possibility that similar types of sounds (e.g. branch cracking), which are hard to tell apart for human listeners, may have been falsely interpreted as a gunshot, and consequently falsely annotated.

2.3. Data labelling

To extract non-gunshot recordings, hereafter referred to as background sounds, we took a random sample of 10,000 recordings each from Sites 1 and 2. We manually checked for the absence of gunshots in these recordings and removed any recordings of human voice.

Due to the low proportion of gunshots, this approach of random sampling and annotation gave too few gunshot examples for model training. We therefore developed and used a deterministic filter to produce a smaller subsample of potential gunshot recordings that could be manually annotated. This filter identified potential gunshots based on the decay of sound pressure over time, and the sound pressure of each frequency band (S1). The filtered files were then classified manually, through a combination of visual inspection of spectrograms and listening to the audio, to separate out the gunshot recordings (Table 1). We added the false positives from this filter to the background training data, to increase the potential for the CNN to learn background sounds that have similar properties to gunshots. We tested the performance of this filter on the ground truthed gunshots and found that it correctly identified 57% of the gunshot recordings. Although use of this filter biases the type of gunshot recordings that we trained our model with, we supplemented these annotated gunshots with a further 357 files that were recorded and labelled during the ground-truthing exercise, which should ameliorate this bias in the dataset (Table 1).

Table 1
Summary of data collected from each site, with the addition of gunshot recordings from a ground truthing exercise. Data from Site 1 and 2 were subsampled and annotated for model training and validation, while data from Site 3 were not annotated, and were used for testing the workflow on a full raw dataset from a survey.

Site	Dataset	Total files	Gunshot	Background
Site 1	Train / Val	2,112,924	325	23,911
Site 2	Train / Val	235,531	67	11,404
Site 3	Test	168,959		
Controlled gunshots	Train / Val	–	357	
		Total	749	35,315

2.4. Data partitioning

The training and validation datasets comprised data collected in Sites 1 and 2, in addition to the ground truthed gunshots. For each of these datasets, we ordered the recordings from each category (background and gunshot) according to the time the recording was made. Recordings made in the first 80% of the sampling period for each category at each site comprised the training dataset, and the remaining recordings, from the last 20% of the sampling period, comprised the validation dataset. The training data were used for model training, whilst the validation data were used for fine-tuning of model hyperparameters, to identify the optimal length of training and diagnose overfitting issues, and to define the decision threshold (above which predictions are counted as gunshots) used for model predictions on the test data. Our test dataset comprised the full, unannotated data collected in Site 3. This dataset was too large to annotate manually, however it was necessary to include the entire dataset to gauge the precision of our classifier when faced with a greater proportion of background sounds, and its generalizability to spatially distinct sites.

2.5. Data balancing

Canonical machine learning algorithms assume an equal distribution between each class, and in cases of highly imbalanced data, the algorithm will be biased towards the majority class (Krawczyk, 2016). We balanced our training dataset to a ratio of 1:1 (gunshots: background) by oversampling the gunshot files to 28,195 to match the number of background sounds in the training data. Additionally, to evaluate the effect of increasing the amount of training background sounds on model performance, we created a second training dataset where background sounds were undersampled to 600 to match the original number of gunshots. We did not balance the validation dataset as we wanted to test how the model performs on a dataset with a more realistic distribution of gunshots. The validation set comprised 150 gunshots and 3995 background sounds.

2.6. Convolutional neural network training

CNNs are a class of neural networks designed to process data that come in the form of grids, such as a 2-D array of pixels that forms a digital image (LeCun et al., 2015). We trained our classifier in Python using OpenSoundscape version 0.5 (Kitzes et al., 2020). Our model consisted of the ResNet18 CNN architecture with a binary classifier output. ResNet18 consists of 17 hidden layers, and one fully connected classification layer. We used initial weights pretrained on over 1.24 million images from 1000 categories from the ImageNet database. Even though these images were not related to our dataset, this approach allowed our model to adapt more quickly to identifying spectrograms. Because of the substantial difference of the images from our training set, we trained the entire model on our training dataset without freezing any layers.

Training the CNN required preprocessing the training and validation data into tensors. A tensor is a multidimensional array of numbers used as the common structure for input data. Our preprocessing pipeline involved creating spectrogram images from the audio clips, applying augmentations, and converting the image to a PyTorch Tensor for model training. We created spectrograms using a window length of 256 and overlap of 128. Spectrograms were bandpassed to include only the frequency bands between 0 and 2000 Hz.

We included augmentations in the pre-processing pipeline to increase the generalisability of our model (Fig. 2; Dufourq et al., 2021; Mushtaq et al., 2021; Sandfort et al., 2019). Augmentations included in the pipeline comprised: i) image overlay, which blends a randomly selected background sample to each sample in the training dataset with an overlay weight of 0.4; ii) colour jitter, which randomly changes the brightness, contrast, saturation, and hue of the image; iii) time mask,

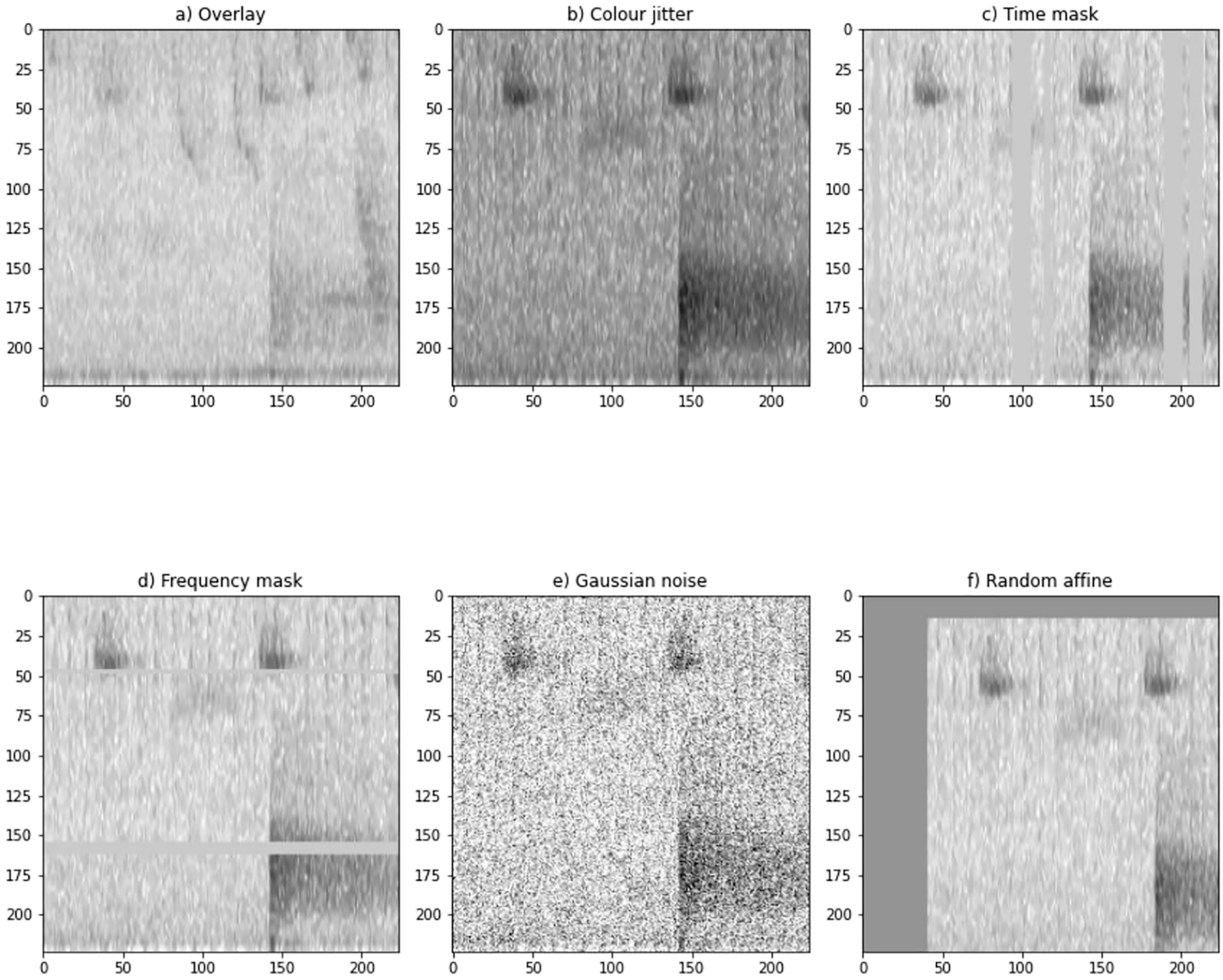

Fig. 2. Examples of the PyTorch tensors created after application of augmentations from the preprocessing pipeline: a) overlay of random background sound with a blending weight of 0.4, b) random change to brightness, contrast, saturation, and hue of image, c) time mask – random application of up to 20 vertical bars to the image, d) frequency mask – random application of up to 20 horizontal bars to the image, e) addition of Gaussian noise with a standard deviation of 0.2, and f) random affine transformation to the image.

which adds up to 20 vertical bars over the image to mask random time bands; iv) frequency mask, which adds up to 20 horizontal bars over the image to mask random frequency bands; v) addition of Gaussian noise to the tensor, with standard deviation of 0.2.; and vi) random affine transformation to the image. We limited the augmentation pipeline for training data only, while the preprocessing pipeline for the validation and test data included no augmentations. We also trained the model without augmentations to evaluate the effect of augmentation on model performance.

We optimised the model using mini-batch gradient descent, with a batch size of 64 samples (Ruder, 2016). We used the default model training parameters provided by OpenSoundscape. We trained the model for 50 epochs (iterations) and selected the epoch with the highest F1 score (the harmonic mean of precision and recall) on the validation data.

2.7. Threshold moving

The model produced real-number numerical scores reflecting confidence of gunshot presence and absence for each sample. These scores were transformed to [0, 1] using the softmax activation function. The scores were subsequently transformed into a binary class label using a decision threshold, above which the class was assigned as a gunshot.

The choice of decision threshold impacts the performance of the classifier due to the trade-off between recall and precision (Knight and Bayne, 2019). A high score threshold will maximise precision, but will have lower recall, and therefore may be more appropriate for ubiquitous and commonly repeated target sounds (e.g. a frequently vocalising frog species Lapp et al. 2021). In contrast, for detecting rare sound events, such as gunshots or vocalisations of a rare species, a lower threshold is more suitable, which maximises recall at the cost of decreased precision. In this case the classifier may be used to guide manual verification of recordings, known as classifier guided listening.

We chose a decision threshold that maximised recall of gunshot events from the validation dataset. We selected a minimum recall of 0.95 to identify the decision threshold score for each of our models.

2.8. Model validation

We evaluated the performance of the models by performing predictions on the validation dataset. The scores were converted into binary predictions using the decision threshold that coincided with 95% recall for each model (Fig. 3). We evaluated model performance by comparing these labels with true class labels.

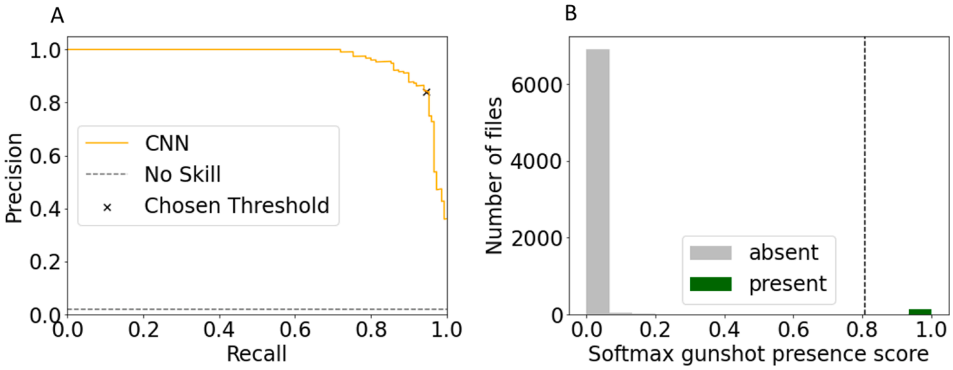

Fig. 3. A) Precision and recall scores of predictions on validation dataset at all possible decision thresholds, with the chosen threshold that coincides with 95% recall highlighted. The no skill line is calculated as the proportion of true positives in the validation dataset. B) Distribution of softmax scores for gunshot presence, obtained from performing predictions on the validation dataset using the decision threshold identified in panel A. True labels of the validation data are represented by the colours of the bars.

2.9. Model implementation

We used our final model to perform predictions on an unseen test dataset from a spatially distinct study site (Site 3). This dataset comprised the full set of recordings obtained from the recording period and was too large to annotate manually. Implementing the model on this larger dataset was essential to gauge the true precision of our model given a much greater source of potential false positive sounds. To evaluate the precision of our model, we performed 'top down' listening, whereby we ranked the prediction scores from highest to lowest, and then listened to all the files that were above the chosen threshold. We manually classified these files as gunshot or background sound to calculate the precision of the classifier (the proportion of true positives out of all the predicted positive files).

3. Results

3.1. Convolutional neural network

The CNN classified gunshots from the validation dataset with high recall (95%) and precision (85%) (Fig. 3). Predictions performed on our test data confirmed that our model generalised well to data from a novel study site and was able to classify gunshots with relatively high precision (57%) despite much greater volumes of potential false positives in this more extensive dataset. Although we could not calculate the recall of the CNN on this unannotated dataset, we assumed it was approximately 95%, as we used the same 95% decision threshold from the validation dataset.

Comparison of models trained with different training datasets and levels of data augmentation showed that the combination of extensive background sounds in the training data along with data augmentation substantially improved model performance (Table 2). Evaluating each of these models using a threshold that coincided with 95% recall showed that the precision of the model was greatly improved under these conditions (Table 2).

A more accurate estimate of precision and model generalisability was estimated from performing predictions followed by top down listening on a full dataset obtained from one month of recordings in a distinct study site (Site 3), situated over 60 km away from the study sites used for the training data. Predictions performed on this dataset of over 160,000 files produced 35 files with scores above the threshold. Listening to these files confirmed that 20/35 were true gunshot events, giving precision of 57%.

Table 2
Comparison of performance of CNNs trained with different training datasets. Metrics refer to predictions performed on the validation dataset at the chosen threshold for each model that coincides with 95% recall. Model a was trained with a small subset of background sounds, and no data augmentation, Model b was trained with the same training set as Model a, but with the addition of data augmentation in the training pipeline. Model c was trained with the entire dataset of background sounds with no augmentation, while the Final model, which is used throughout this study was trained with the entire dataset of background sounds and data augmentation.

	Model a	Model b	Model c	Final model
Augmentation	No	Yes	No	Yes
Training background sounds	600	600	28,195	28,195
Recall	0.95	0.95	0.95	0.95
Precision	0.05	0.05	0.10	0.85
False positives	2952	2759	1266	26

3.2. Combined pipeline

We calculated the overall recall of our combined pipeline starting from the on-board detector through to the CNN by taking the product of recall from each stage. Recall of the on-board detector was estimated at 0.84, while recall of the CNN was estimated at 0.95, giving an overall recall of 0.80.

4. Discussion

Unsustainable hunting is one of the foremost threats to global biodiversity, yet we have a scant understanding of it on a spatial scale due to our inability to monitor this cryptic activity. Our results demonstrate the utility of deep learning for automated gunshot detection from acoustic data collected in tropical forests. Our CNN classifier achieved high recall (0.95) of gunshots on the validation data with a precision of 0.85. We implemented the pipeline on a full dataset of recordings collected at a spatially distinct site and found that gunshots were identified with a precision of 0.57 and estimated recall of 0.80. Crucially our approach maximises recall of rare gunshot events, whilst reducing the number of the possible gunshot files to a more manageable dataset for human verification.

The performance of this approach, including the maximum overall recall that can be achieved, is constrained by the on-board gunshot detection algorithm implemented on the AudioMoths. Survey design decisions such as recording duration, overall survey length, and use of on-board detectors are highly dependent on the application and behaviour of the target species. For ecological research, on-board detection algorithms are often not ideal, as they discard data that may be important for future research. However, for use by conservation managers with specific goals and limited data management resources, they may be the most practical route forward as their use can increase the overall survey duration by reducing power consumption and storage used on SD cards, while reducing overall data storage and management demands. In this case, the on-board detector allowed 24 h listening for gunshots for a manyfold longer overall duration than would be possible with 24 h continuous recording (Prince et al., 2019).

Comparison of models trained with different training datasets showed that incorporation of a wide variety of background sounds from our study environment in the training dataset substantially improved model performance (Bergler et al., 2019; Florentin et al., 2020). Our annotated dataset was collected over 14 months at 25 sensor locations in Belize. Although this is not considered a particularly large dataset by machine learning standards, crucially it captured the highly heterogeneous and noisy soundscapes characteristic to tropical forests. Our realistic, noisy field data provided both realistic gunshot recordings and a wide range of potential false positive sounds. Given how rare gunshot events are in comparison to these potential false positives, the ability of the classifier to learn the gunshot-like (prone to false positive) sounds occurring within a forest environment was critical, as any learning in this domain can substantially reduce the amount of false positive predictions.

Data augmentation techniques were also key to improving the performance of our model. CNNs are sensitive to overfitting on training data, as they are narrowly focussed on the annotated training dataset. To avoid this issue, CNNs require extremely large annotated datasets, that are typically unachievable in an ecological monitoring scenario. Consequently the transferability of models to novel contexts is often an issue (Gibb et al., 2019). The use of data augmentation techniques can address this issue by introducing novel data to the model, thereby reducing the chance of overfitting on the training dataset. In our case, augmentation substantially improved the precision of our classifier when tested on the validation dataset.

Implementation of our classification model on a full dataset collected at a novel study site over 60 km away from the training and validation data sites indicated good generalizability of our model. Predictions performed on this dataset of over 160,000 sound clips filtered out the vast majority of the negative clips, leaving just 35 sounds for manual review. Out of these predicted positives, 20 were identified as gunshots, giving our model a precision of 57%. Although true recall cannot be measured using this approach, we are comfortable with the assumption that the 95% recall from our validation data (and 80 % recall of the combined pipeline) should transfer to this new dataset. The reduction in precision of the classifier from the validation to the test dataset was expected, as the validation dataset had a smaller distribution of negative files compared to the test dataset (96.38 % background sounds in the validation data compared to ~ 99.99 % background sounds in the test data), which results in increased overlap of classification scores for positive and negative classes. Although our classifier generalised well to this spatially distinct study site, the generalisability to more spatially distant tropical forests is uncertain, as is the model's generalisability to different types of landscapes and firearms type. Use of this approach in novel regions would require testing the pipeline on data collected in the new area, quantifying recall and precision, and potentially retraining the model with additional data.

Our workflow has demonstrated the inherent challenges associated with robust acoustic detection of gunshots. The main challenge with gunshot detection is the fact that the acoustic signal is simple and nondescript and is easily confused with background sounds such as branch cracks, knocks, and mechanical sounds, by both human listeners and detection algorithms. Just as it is legitimately difficult for humans to distinguish these sounds using both listening and spectrogram inspection, it follows that the CNN will also struggle. Target sounds that are similarly simple and nondescript include woodpecker drums and cetacean clicks, and studies have similarly struggled to train a spectrogram based CNN to robustly detect these sounds without including a substantial number of false positives (Bergler et al., 2019; Florentin et al., 2020). In contrast, complex bird songs may be more easily learned by CNNs as they produce more distinctive spectrogram images (Florentin et al., 2020). Despite achieving lower performance metrics on these simple sounds, these studies still found value in using spectrogram based CNNs to reduce the vast datasets into more manageable, tentatively annotated datasets of potential positives that can be manually reviewed (Florentin et al., 2020). Likewise, application of our detection workflow could make wide-scale gunshot monitoring in forests a more feasible option.

To our knowledge, just two other studies have applied automated detection for acoustic monitoring of gun hunting in tropical forests (Dobbins et al., 2020; Wrege et al., 2017). Wrege et al. (2017) implemented a template cross-correlation method to detect gunshots from field data collected in Central African tropical forests, using a template that included nine example gunshots. While this approach achieved high recall, it resulted in a high level of false positive predictions, and the precision was as low as 0.0003. Classifiers with such low precision would be inappropriate for scaling up to larger sensor deployments, as they would require inordinate amounts of time dedicated to manual classification of the false positive predictions. Dobbins et al. (2020) took a different approach, involving clustering and machine learning using Kaleidoscope 5 software. As the details of this methodology, including metrics of precision and recall were not reported, it is not possible to compare this method to ours, or for other users to reproduce the study.

An alternative technique has recently been proposed, for unsupervised deep learning of anomalous sounds, including gunshots and chainsaws (Sethi et al., 2020b). Instead of classifying target sounds, this method passes sounds through a CNN and removes the final classification layer, producing a set of numerical features for each clip. Anomalous sounds are then identified by fitting a probability distribution to these features. Although this anomaly-detection method has potential for generalisability across a variety of different landscapes (Sethi et al., 2020b), as yet we know of no reported performance metrics for this approach in terms of false positive rate and how many different sound classes are classified as anomalies. As this method is less targeted than supervised approaches, it likely sacrifices accuracy for generalizability. Gunshots were only reliably identified using this approach at distances of 25 m, while further gunshots could not be distinguished from background sound (Sethi et al., 2020b). This highly constrained detection distance would miss many gunshots from sensors such as AudioMoths, which record gunshots from up to 1 km away in tropical forests (Prince et al., 2019). The choice of classification method ultimately reflects the objectives of the study and the availability of annotated training data; unsupervised techniques such as this prioritise generalisability over detection of specific target sounds and may be more useful for exploratory analysis of soundscapes across wide extents, whereas supervised classifiers are more appropriate for maximising the detection of a specific target sound in a chosen study system.

The purpose of our approach is to provide conservation managers with a long term retrospective dataset of hunting pressure, which can be used to assess the effectiveness of their current strategies and to plan more effective future patrols. Key information from this dataset includes i) temporal trends in hunting activity, and ii) spatial hotspots of hunting activity. This adaptive management approach (Astaras et al., 2020; Hötte et al., 2016) contrasts to other systems that provide real time alerts with the aim of allowing managers to act quickly on live data (O'Donoghue and Rutz, 2016; Sarma and Baruah, 2015). In cases such as ours, where the protected areas are rugged, remote, difficult to access, and managers have limited resources, we believe this retrospective approach is more beneficial than a system of real time detections. From our experience of working in this region, rangers would unlikely be able to act on real time alerts of gunshots due to limited personnel and difficulty in physically getting to each location. As patrols in this region require careful advanced planning, information to guide this process would be of greater value than real time alerts. Furthermore, the concept of real-time alerts brings to light two additional issues. Firstly, it is evident from our study and others, that highly precise detection of gunshots (without substantially sacrificing recall) remains a major challenge, even in urban areas with far greater resources. For example, the use of 'Shotspotter' acoustic gunshot detection technology by the Chicago Police Department found that under 10% of gunshot alerts were associated with a likely gunshot event, resulting in over 37,000 police department responses in one year to locations with no evidence of gun crime (Ferguson and Witzburg, 2021). Consequently, a real time detection system would likely provide too many false alarms to be practical. Secondly, real-time alerts present substantial social and ethical issues around compromising the personal safety and wellbeing of rangers and poachers alike (Sandbrook, 2015; Simlai and Sandbrook, 2021). Ultimately the case of gunshot detection highlights the need to reflect on the objective of each monitoring scenario to appropriately optimise the trade-off between factors such as recall and precision, and generalisability and accuracy.

5. Conclusion

Acoustic technology offers a multitude of opportunities for monitoring biodiversity, environmental health, and human disturbance such as gun hunting. However, there is currently a mismatch between the speed that affordable hardware is being developed and the ability of ecologists to process the vast amounts of data collected. Prerequisites for bridging this gap include case studies that utilise open access software and provide fully reproducible workflows. An additional barrier is the lack of public audio datasets obtained from the field, especially in tropical regions, which can be used for model training or benchmarks for comparative tests of algorithms. In the case of gunshot detection in the tropics, to our knowledge there are no publicly archived datasets, which is a substantial limitation to development of useful detectors. Lifting these barriers would hasten development of detection algorithms with field utility, and ultimately would allow the development of open access and more user-friendly software that is accessible to ecologists.

CRediT authorship contribution statement

Lydia K.D. Katsis: Conceptualization, Methodology, Writing – original draft, Software, Investigation. **Andrew P. Hill:** Investigation, Methodology. **Evelyn Piña-Covarrubias:** Investigation. **Peter Prince:** Data curation, Investigation, Methodology. **Alex Rogers:** Supervision. **C. Patrick Doncaster:** Investigation, Software, Supervision, Writing – review & editing. **Jake L. Snaddon:** Investigation, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Natural Environmental Research Council [grant number NE/S007210/1], the Royal Geographical Society Dudley Stamp Memorial Award, and an Idea Wild equipment grant, awarded to L.K.D. Katsis. Additional support included an Interdisciplinary Science and Enterprise grant from the University of Southampton Institute for Life Sciences, an Engineering and Physical Sciences Research Council (EPSRC) Impact Acceleration Account award from the University of Oxford, an EPSRC Studentship (1658469) to A. Hill, a Natural Environmental Research Council SPITFIRE DTP award (NE/L002531/1) to P. Prince, and a CONACyT Studentship (202650) and Rufford Grant (17047-1) to E. Piña Covarrubias. We thank the following individuals and organisations for their indispensable help in the field: Ray and Vicki Snaddon and the lodge staff at Pook’s Hill Lodge; Panthera Belize, including Emma Sanchez, Shirley Chang, Rigoberto Saqui, Mitch Brakeman, Rebecca Foster, and Bart Harmsen; Ya’axché Conservation Trust, including Said Gutierrez, Oscar Requena, Anignazio Makin, Andres Chen, Gerasimo Coc, Rosendo Coy, and the Ya’axché Enforcement Team. We thank Dr. Justin Kitzes, Sam Lapp, Tessa Rhineheart, R. Patrick Lyon, and Lauren Chronister for their invaluable comments and suggestions on this manuscript and for providing computing resources from their lab at the University of Pittsburgh. We thank the two anonymous reviewers for useful feedback and suggestions.

Data availability

All recordings and annotations used for training and validating the gunshot detection CNN are available on Mendeley Data (DOI: 10.17632/x48cwz364j.3, https://data.mendeley.com/datasets/x48cwz364j/3) and all Python code underlying the gunshot detection CNN, including the trained classifier, are available on GitHub (https://github.com/lydiakatsis/tropical_forest_gunshot_classifier).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ecolind.2022.109128.

References

Astaras, C., Linder, J., Wrege, P., Orume, R., Johnson, P., Macdonald, D., 2020. Boots on the ground: the role of passive acoustic monitoring in evaluating anti-poaching patrols. Environ. Conserv. 47, 213–216. https://doi.org/10.1017/S0376892920000193.

Astaras, C., Linder, J.M., Wrege, P., Orume, R.D., Macdonald, D.W., 2017. Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests. Front. Ecol. Environ. 15, 233–234. https://doi.org/10.1002/fee.1495.

Bajzik, J., Prinosil, J., Koniar, D., 2020. Gunshot Detection Using Convolutional Neural Networks, 2020 24th International Conference Electronics, pp. 1-5, 10.1109/IEEECONF49502.2020.9141621.

Barnosky, A.D., Matzke, N., Tomiya, S., Wogan, G.O., Swartz, B., Quental, T.B., Marshall, C., McGuire, J.L., Lindsey, E.L., Maguire, K.C., 2011. Has the Earth’s sixth mass extinction already arrived? Nature 471, 51.

Benítez-López, A., Santini, L., Schipper, A.M., Busana, M., Huijbregts, M.A., 2019. Intact but empty forests? Patterns of hunting-induced mammal defaunation in the tropics. PLoS Biol. 17, e3000247.

Bergler, C., Schröter, H., Cheng, R.X., Barth, V., Weber, M., Nöth, E., Hofer, H., Maier, A., 2019. ORCA-SPOT: An Automatic Killer Whale Sound Detection Toolkit Using Deep Learning. Scientific Reports 9, 10997. https://doi.org/10.1038/s41598-019-47335-w.

Campos-Cerqueira, M., Aide, T.M., 2016. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. Methods Ecol. Evol. 7, 1340–1348. https://doi.org/10.1111/2041-210X.12599.

Ceballos, G., Ehrlich, P.R., Raven, P.H., 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. Proc. Natl. Acad. Sci. U.S.A. 117, 13596–13602. https://doi.org/10.1073/pnas.1922686117.

Chacon-Rodriguez, A., Julian, P., Castro, L., Alvarado, P., Hernandez, N., 2011. Evaluation of gunshot detection algorithms. IEEE Trans. Circuits Syst. I-Regul. Pap. 58, 363–373. https://doi.org/10.1109/tcsi.2010.2072052.

Critchlow, R., Plumptre, A.J., Alidria, B., Nsubuga, M., Driciru, M., Rwetsiba, A., Wanyama, F., Beale, C.M., 2017. Improving law-enforcement effectiveness and efficiency in protected areas using ranger-collected monitoring data. Conserv. Lett. 10, 572–580. https://doi.org/10.1111/conl.12288.

Critchlow, R., Plumptre, A.J., Driciru, M., Rwetsiba, A., Stokes, E.J., Tumwesigye, C., Wanyama, F., Beale, C.M., 2015. Spatiotemporal trends of illegal activities from ranger-collected data in a Ugandan national park. Conserv. Biol. 29, 1458–1470. https://doi.org/10.1111/cobi.12538.

Deith, M.C.M., Brodie, J.F., 2020. Predicting defaunation: accurately mapping bushmeat hunting pressure over large areas. Proc. Royal Society B: Biol. Sci. 287, 20192677. https://doi.org/10.1098/rspb.2019.2677.

Do Nascimento, L.A., 2020. Ecoacoustic Methods for Multi-Taxa Animal Surveys in the Amazon.

Dobbins, M., Sollmann, R., Menke, S., Almeyda Zambrano, A., Broadbent, E., 2020. An integrated approach to measure hunting intensity and assess its impacts on mammal populations. J. Appl. Ecol. https://doi.org/10.1111/1365-2664.13750.

Dobson, A., Milner-Gulland, E., Aebischer, N.J., Beale, C.M., Brozovic, R., Coals, P., Critchlow, R., Dancer, A., Greve, M., Hinsley, A., 2020. Making messy data work for conservation. One Earth 2, 455–465. https://doi.org/10.1016/j.oneear.2020.04.012.

Dobson, A.D.M., Milner-Gulland, E.J., Beale, C.M., Ibbett, H., Keane, A., 2019. Detecting deterrence from patrol data. Conserv Biol 33, 665–675. https://doi.org/10.1111/cobi.13222.

Dufourq, E., Durbach, I., Hansford, J.P., Hoepfner, A., Ma, H.D., Bryant, J.V., Stender, C. S., Li, W.Y., Liu, Z.W., Chen, Q., Zhou, Z.L., Turvey, S.T., 2021. Automated detection of Hainan gibbon calls for passive acoustic monitoring. Remote Sens. Ecol. Conserv. 10.1002/rse2.201, 10.1002/rse2.201.

Ferguson, J.M., Witzburg, D., 2021. The Chicago Police Department’s Use of Shotspotter Technology, City of Chicago Office of Inspector General.

Ferreguetti, Á.C., Pereira-Ribeiro, J., Prevedello, J.A., Tomás, W.M., Rocha, C.F.D., Bergallo, H.G., 2018. One step ahead to predict potential poaching hotspots: Modeling occupancy and detectability of poachers in a neotropical rainforest. Biol. Conserv. 227, 133–140. https://doi.org/10.1016/j.biocon.2018.09.009.

Florentin, J., Dutoit, T., Verlinden, O., 2020. Detection and identification of European woodpeckers with deep convolutional neural networks. Ecol. Informatics 55, 101023. https://doi.org/10.1016/j.ecoinf.2019.101023.

Foster, R.J., Harmsen, B.J., Macdonald, D.W., Collins, J., Urbina, Y., Garcia, R., Doncaster, C.P., 2016. Wild meat: a shared resource amongst people and predators. Oryx 50, 63–75. https://doi.org/10.1017/s003060531400060x.

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol. Evol. 10, 169–185. https://doi.org/10.1111/2041-210X.13101.

Hansen, M.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A., Tyukavina, A., Thau, D., Stehman, S.V., Goetz, S.J., Loveland, T.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O., Townshend, J.R.G., 2013. High-resolution global maps of 21st-century forest cover change. Science 342, 850–853. https://doi.org/10.1126/science.1244693.

Hill, A.P., Prince, P., Snaddon, J.L., Doncaster, C.P., Rogers, A., 2019. AudioMoth: a low-cost acoustic device for monitoring biodiversity and the environment. HardwareX 6, e00073.

Hossain, A.N.M., Barlow, A., Barlow, C.G., Lynam, A.J., Chakma, S., Savini, T., 2016. Assessing the efficacy of camera trapping as a tool for increasing detection rates of wildlife crime in tropical protected areas. Biol. Conserv. 201, 314–319. https://doi.org/10.1016/j.biocon.2016.07.023.

Hötte, M.H.H., Kolodin, I.A., Bereznuk, S.L., Slaght, J.C., Kerley, L.L., Soutyrina, S.V., Salkina, G.P., Zaumyslova, O.Y., Stokes, E.J., Miquelle, D.G., 2016. Indicators of success for smart law enforcement in protected areas: a case study for Russian Amur tiger (Panthera tigris altaica) reserves. Integr. Zool. 11, 2–15. https://doi.org/10.1111/1749-4877.12168.

Hrabina, M., Sigmund, M., Ieee, 2016. Implementation of Developed Gunshot Detection Algorithm on TMS320C6713 Processor. Ieee, New York, 10.1109/SAI.2016.7556087.

IPBES, 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services., in: S. Brondizio, J. Settele, S. Díaz, Ngo, H.T. (Eds.), IPBES secretariat, Bonn, Germany.

Jones, K.R., Venter, O., Fuller, R.A., Allan, J.R., Maxwell, S.L., Negret, P.J., Watson, J.E. M., 2018. One-third of global protected land is under intense human pressure. Science 360, 788–791. https://doi.org/10.1126/science.aap9565.

Kahl, S., Wood, C.M., Eibl, M., Klinck, H., 2021. BirdNET: A deep learning solution for avian diversity monitoring. Ecol. Informatics 61, 101236. https://doi.org/10.1016/j.ecoinf.2021.101236.

Khunarsa, P., Lursinsap, C., Raicharoen, T., 2010. Impulsive environment sound detection by neural classification of spectrogram and mel-frequency coefficient images, Advances in Neural Network Research and Applications. Springer 337–346.

Kitzes, J., Moore, B., Rhinehart, T.A., Lapp, S., 2020. OpenSoundscape.org.

Knight, E.C., Bayne, E.M., 2019. Classification threshold and training data affect the quality and utility of focal species data processed with automated audio-recognition software. Bioacoustics 28, 539–554. https://doi.org/10.1080/09524622.2018.1503971.

Krawczyk, B., 2016. Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence 5, 221–232. https://doi.org/10.1007/s13748-016-0094-0.

Lapp, S., Wu, T., Richards-Zawacki, C., Voyles, J., Rodriguez, K.M., Shamon, H., Kitzes, J., 2021. Automated detection of frog calls and choruses by pulse repetition rate. Conserv. Biol. https://doi.org/10.1111/cobi.13718.

Laurance, W.F., Useche, D.C., Rendeiro, J., Kalka, M., Bradshaw, C.J., Sloan, S.P., Laurance, S.G., Campbell, M., Abernethy, K., Alvarez, P., 2012. Averting biodiversity collapse in tropical forest protected areas. Nature 489, 290–294. https://doi.org/10.1038/nature11318.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444. https://doi.org/10.1038/nature14539.

Liu, Y., Cheng, Z., Liu, J., Yassin, B., Nan, Z., Luo, J., 2019. AI for Earth: Rainforest Conservation by Acoustic Surveillance. arXiv preprint arXiv:1908.07517.

Morehead, A., Ogden, L., Magee, G., Hosler, R., White, B., Mohler, G., 2019. Low Cost Gunshot Detection using Deep Learning on the Raspberry Pi. In: 2019 IEEE International Conference on Big Data (Big Data), pp. 3038–3044. https://doi.org/10.1109/BigData47090.2019.9006456.

Mushtaq, Z., Su, S.-F., 2020. Environmental sound classification using a regularized deep convolutional neural network with data augmentation. Appl. Acoustics 167, 107389. https://doi.org/10.1016/j.apacoust.2020.107389.

Mushtaq, Z., Su, S.-F., Tran, Q.-V., 2021. Spectral images based environmental sound classification using CNN with meaningful data augmentation. Appl. Acoustics 172, 107581. https://doi.org/10.1016/j.apacoust.2020.107581.

Nimmy, P., Rajesh, K.R., Nimmy, M., Vishnu, S., Ieee, 2018. Shock Wave and Muzzle Blast Identification Techniques Utilizing Temporal and Spectral Aspects of Gunshot Signal. Ieee, New York, 10.1109/RAICS.2018.8635092.

O'Donoghue, P., Rutz, C., 2016. Real-time anti-poaching tags could help prevent imminent species extinctions. J. Appl. Ecol. 53, 5–10. https://doi.org/10.1111/1365-2664.12452.

Peres, C.A., 2001. Synergistic effects of subsistence hunting and habitat fragmentation on Amazonian forest vertebrates. Conserv. Biol. 15, 1490–1505. https://doi.org/10.1046/j.1523-1739.2001.01089.x.

Peres, C.A., Barlow, J., Laurance, W.F., 2006. Detecting anthropogenic disturbance in tropical forests. Trends Ecol. Evol. 21, 227–229. https://doi.org/10.1016/j.tree.2006.03.007.

Picciulin, M., Kéver, L., Parmentier, E., Bolgan, M., 2019. Listening to the unseen: passive acoustic monitoring reveals the presence of a cryptic fish species. Aquatic Conservation: Marine and Freshwater Ecosystems 29, 202–210. https://doi.org/10.1002/aqc.2973.

Pijanowski, B.C., Farina, A., Gage, S.H., Dumyahn, S.L., Krause, B.L., 2011. What is soundscape ecology? An introduction and overview of an emerging new science. Landscape Ecol. 26, 1213–1232. https://doi.org/10.1007/s10980-011-9600-8.

Plumptre, A.J., Fuller, R.A., Rwetsiba, A., Wanyama, F., Kujirakwinja, D., Driciru, M., Nangendo, G., Watson, J.E., Possingham, H.P., 2014. Efficiently targeting resources to deter illegal activities in protected areas. J. Appl. Ecol. 51, 714–725. https://doi.org/10.1111/1365-2664.12227.

Prince, P., Hill, A., Covarrubias, E.P., Doncaster, P., Snaddon, J.L., Rogers, A., 2019. Deploying acoustic detection algorithms on low-cost, open-source acoustic sensors for environmental monitoring. Sensors 19, 23. https://doi.org/10.3390/s19030553.

Priyadarshani, N., Marsland, S., Castro, I., 2018. Automated birdsong recognition in complex acoustic environments: a review. J. Avian Biol. 49, jav-01447. https://doi.org/10.1111/jav.01447.

Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.

Ruff, Z.J., Lesmeister, D.B., Appel, C.L., Sullivan, C.M., 2021. Workflow and convolutional neural network for automated identification of animal sounds. Ecological Indicators 124, 107419. https://doi.org/10.1016/j.ecolind.2021.107419.

Sandbrook, C., 2015. The social implications of using drones for biodiversity conservation. Ambio 44, 636–647. https://doi.org/10.1007/s13280-015-0714-0.

Sandfort, V., Yan, K., Pickhardt, P.J., Summers, R.M., 2019. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. Scientific Reports 9, 16884. https://doi.org/10.1038/s41598-019-52737-x.

Sarma, T., Baruah, V., 2015. Real time poaching detection: A design approach, 2015 International Conference on Industrial Instrumentation and Control (ICIC). IEEE, pp. 922-924.

Sethi, S.S., Ewers, R.M., Jones, N.S., Signorelli, A., Picinali, L., Orme, C.D.L., 2020a. SAFE Acoustics: an open-source, real-time eco-acoustic monitoring network in the tropical rainforests of Borneo. bioRxiv 10.1111/2041-210X.13438.

Sethi, S.S., Jones, N.S., Fulcher, B.D., Picinali, L., Clink, D.J., Klinck, H., Orme, C.D.L., Wrege, P.H., Ewers, R.M., 2020b. Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set. Proceedings of the National Academy of Sciences, 202004702, 10.1073/pnas.2004702117.

Simlai, T., Sandbrook, C., 2021. Digital surveillance technologies in conservation and their social implications. Conserv. Technol. 239 https://doi.org/10.1111/csp2.374.

Singh, V., Ray, K.C., Tripathy, S., 2020. Robust Gunshot Features and Its Classification Using Support Vector Machine for Wildlife Protection, Electronic Systems and Intelligent Computing. Springer, pp. 939-948, 10.1007/978-981-15-7031-5_89.

Sugai, L.S.M., Silva, T.S.F., Ribeiro Jr, J.W., Llusia, D., 2018. Terrestrial passive acoustic monitoring: review and perspectives. BioScience 69, 15–25. https://doi.org/10.1093/biosci/biy147.

Valenzise, G., Gerosa, L., Tagliasacchi, M., Antonacci, E., Sarti, A., Ieee, 2007. Scream and gunshot detection and localization for audio-surveillance systems. Ieee, New York, 10.1109/avss.2007.4425280.

Van der Merwe, J., Jordaan, J., 2013. Comparison between general cross correlation and a template-matching scheme in the application of acoustic gunshot detection, 2013 Africon. IEEE 1–5. https://doi.org/10.1109/AFRCON.2013.6757698.

Wrege, P.H., Rowland, E.D., Keen, S., Shiu, Y., 2017. Acoustic monitoring for conservation in tropical forests: examples from forest elephants. Methods Ecol. Evol. 8, 1292–1301. https://doi.org/10.1111/2041-210X.12730.

WWF, 2016. Living planet report: risk and resilience in a new era, WWF International, 978-2-940529-40-7.

Ziegler, S., Fa, J.E., Wohlfart, C., Streit, B., Jacob, S., Wegmann, M., 2016. Mapping bushmeat hunting pressure in Central Africa. Biotropica 48, 405–412. https://doi.org/10.1111/btp.12286.