IMPERIAL COLLEGE LONDON

THESIS

Using acoustic monitoring and machine learning to detect illegal hunting in Costa Rica

Author:
Jacob GRIFFITHS

Supervisors:
Dr. Cristina BANKS-LEITE
Dr. James ROSINDELL

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science at Imperial College London

Formatted in the journal style of

Submitted for the MSc in Computational Methods in Ecology and Evolution

Department of Life Sciences

August 28, 2019

Declaration of Authorship

I, Jacob GRIFFITHS, declare that this thesis titled, "Using acoustic monitoring and machine learning to detect illegal hunting in
Costa Rica" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

IMPERIAL COLLEGE LONDON

Abstract

Faculty of Natural Sciences
Department of Life Sciences

Master of Science

Using acoustic monitoring and machine learning to detect illegal hunting in Costa Rica

by Jacob GRIFFITHS

The rapid loss of global biodiversity makes the development of more powerful and cost-effective mitigation strategies of ever-increasing importance. Advances in passive audio monitoring (PAM) and machine learning are providing some of those vital strategies. The reduction in costs and the increasing storage capacity of acoustic sensors are enabling the cultivation of large acoustic datasets which can be utilised for the monitoring of biodiversity, the detection of illegal activity, and soundscaping, among other uses. Despite the potential benefits of such large acoustic datasets, size has its disadvantages, as traditional analytical techniques become far more labour-intensive or completely unusable. This study explores the use of advanced machine learning techniques, specifically deep convolutional neural networks (CNNs), as a potential tool to help identify illegal hunting in protected areas. In this study I demonstrate that, despite the relatively low precision achieved at this stage, CNNs are still a highly effective method in improving the detection of illegal hunting, and that analysis of acoustic data can provide useful ecological and practical information regarding hunting behaviours. I successfully trained a CNN using acoustic data from a National Park in Belize and was able to use that model to detect gunshots in acoustic data from the Osa Peninsula in Costa Rica. I also found that the frequency of gunshots can be predicted by the day of the week and the time of day. Specifically, most gunshots were detected in the morning and far fewer were detected at night, in stark contrast to the findings of similar studies. These findings will not only reduce the labour-intensiveness of current hunting detection, but also provides immediately useful information regarding hunting tendencies. This will allow rangers in protected areas, with their notably limited resources, to develop more effective management strategies and thus aid conservation in the region.

Acknowledgements

Write acknowledgements here

Contents

List of Figures

List of Tables

List of Abbreviations

PAM	**P**assive **A**udio **M**onitoring
MFC	**M**el **F**requency **C**epstrum
MFCCs	**M**el **F**requency **C**epstrum **C**oefficients
CNN	**C**onvolutional **N**eural **N**etwork

Introduction

1.1 Biodiversity loss

It is well documented that global biodiversity loss is accelerating. Anthropogenic factors are often touted as the leading cause, either directly through deforestation and hunting, or indirectly through climate change and the introduction of invasive species (Chiarucci, Bacaro, and Scheiner, 2011; Doherty et al., 2016; Newbold et al., 2015). A meta-analysis by De Vos et al., 2015 gave an arguably conservative estimate of current biodiversity loss as being 1,000 times greater than the background 'natural' rate, with a magnitude of 10,000 times greater also being plausible (Ceballos et al., 2015). This alarming rate of loss not only affects the lost taxa themselves, but can also lead to extensive reduction in ecosystem multifunctionality, typically impacting poorer human communities (Chiarucci, Bacaro, and Scheiner, 2011; Allan et al., 2015; Fanin et al., 2018; Cardinale et al., 2012; Fanin et al., 2018).

Tropical forests are no exception to this global trend. In addition to the adverse effects of deforestation on biodiversity, it is thought that other forms of anthropogenic disturbance, such as vehicles, hunting, and light pollution, can double the biodiversity loss (Barlow et al., 2016).

1.2 Difficulties of measuring biodiversity and its loss

Whilst it is universally agreed that biodiversity is being lost (Cardinale et al., 2012; Ceballos et al., 2015; Fanin et al., 2018), measuring this can be inconsistent and even erroneous (Rocchini et al., 2018) for the following reasons. Firstly, only about 15% of all species have been described, and therefore for the vast majority of species on Earth, there is no census data (Chapman, 2009). Secondly, of those that have been described, very little is known of their distribution, population size, ecology, and life histories, with many species only known by a single specimen (Chapman, 2009). Finally, of those species that have been described and studied in greater depth, there are inconsistencies in survey methods and often a lack of baseline measures to compare to (Society, 2003). The survey method deemed 'best' is often specific to a certain level of organisation and spatial scale of interest, such as satellite imagery and ground surveys for rainforest plant surveys. Despite the challenges and shortcomings associated with measuring biodiversity loss, it is indisputable that its acceleration is rapid, making urgent the development of programmes to assess and monitor biodiversity which are suitable for the answering of large-scale ecological questions (Chiarucci, Bacaro, and Scheiner, 2011). Whilst it is universally agreed that biodiversity is being lost (Cardinale et al., 2012; Ceballos et al., 2015; Fanin et al., 2018), measuring this can be inconsistent and even erroneous (Rocchini et al., 2018) for the following reasons. Firstly, only about 15% of all species have been described, and therefore for the vast majority of species on Earth, there is no census data (Chapman, 2009). Secondly, of those that have been described, very little is known of their distribution, population size, ecology, and life histories, with many species only known by a single specimen (Chapman, 2009). Finally, of those species that have been described and studied in greater depth, there are inconsistencies in survey methods and often a lack of baseline measures to compare to (Society, 2003; Rocchini et al., 2018). The survey method deemed 'best' is often specific to a certain level of organisation and spatial scale of interest, such as satellite imagery and ground surveys for rainforest plant surveys (Rocchini et al., 2018). Despite the challenges and shortcomings associated with measuring biodiversity loss, it is indisputable that its acceleration is rapid, making urgent the development of programmes to assess and monitor biodiversity which are suitable for the answering of large-scale ecological questions (Chiarucci, Bacaro, and Scheiner, 2011; Rocchini et al., 2018).

1.3 Costa Rica and conservation

Costa Rica is no exception to the aforementioned trend of biodiversity loss (Höbinger et al., 2012), with the Osa Peninsula being an area of focus for recent studies as it contains a mixture of protected, partially-protected, and unprotected land, including three national parks, Corcovado, Piedras Blancas and the Terreba-Sierpe wetlands (Lawson, 2019). Whilst protected areas have benefited some species, others, such as the Geoffroy's spider monkey, *Ateles geoffroyi*, are struggling. This is due to their diet, need for mature trees, and need for large areas to roam, with typical home ranges of 4 km^2, which is increasingly limiting their range and may be isolating populations, in turn reducing their survival and genetic variability (Chapman, Chapman, and McLaughlin, 1989). *A. geoffroyi* is classified as endangered by the International Union for Conservation of Nature (IUCN) due to a 50% reduction in numbers over the last 45 years (Cuarón et al., 2008). This reduction may be negatively impacting other species, as *A. geoffroyi* is known to disperse the seeds of up to 150 tree species (Roosmalen, 1985; Pacheco and Simonetti, 2000). As well as habitat fragmentation, *A. geoffroyi* is being subjected to hunting in both protected and non-protected areas. Aquino et al., 2013 found hunted populations of *A. geoffroyi* in Peru were 70-80% less dense than non-hunted populations. Admittedly, the monitoring and prevention of hunting in protected areas is often difficult in large reserves, due to the limited resources available to both rangers and conservationists. However, a recent study by Hill et al., 2018 demonstrated that gunshots can be detected with acoustic sensors up to 1km away from the source, opening up the possibility of a more effective and cheaper migration strategy.

1.4 Passive acoustic monitoring and its advantages

Passive acoustic monitoring (PAM) is becoming an increasingly popular method for large-scale biodiversity monitoring, primarily due to its relatively low cost (Browning et al., 2017; Gibb et al., 2019). This involves deploying sound recorders in an environment and having them record for days or weeks at a time to either track a vocal species directly or to use a vocal species as a proxy for another species or the ecosystem as a whole. Previously, methods such as PAM have been greatly limited by high implementation costs, a lack of digitisation, and low data storage capacity (Merchant et al., 2015). However, advances over the last 10-15 years have reduced the impact of these constraints dramatically (Hill et al., 2018; Gibb et al., 2019). One audio sensor in particular that has been developed recently in a collaborative project between the University of Oxford and University of Southampton, AudioMoth, is making PAM not only a viable option for monitoring biodiversity loss, but also one of the best methods available (Hill et al., 2018).

In addition to the direct monitoring of biodiversity, PAM can also be used to track other acoustics which may be relevant to conservation, such as gunshots, which are generally associated with illegal hunting, particularly in protected areas. Astaras et al., 2017 used PAM in a national park in Cameroon to successfully monitor the rates of hunting in the area. They found that most hunting (68.6%) occurred at night when ranger patrols were minimal, and that there was more illegal activity during the week. Astaras et al., 2017 therefore argue that the hunting is for the illegal meat trade rather than for sustenance or sport, as the meat is gathered during the week for the Saturday market days. The cost of the PAM equipment was recorded by Astaras et al., 2017 as being quite high, which may limit the availability of its implementation in other national parks. However, the recent development of much cheaper audio sensors by Hill et al., 2018 may aid the spread of these techniques in conservation areas around the world. Digitisation has also made PAM a more viable survey mehod as it allows significantly longer recording times and records the data in a more appropriate format for computer analysis (Hill et al., 2018; Gibb et al., 2019).

1.5 Machine learning

Once audio data are collected, ecological information can be extracted manually or automatically. Manual extraction involves either auditory or visual inspection of the data and classification of the sounds, which naturally incurs some bias based on the skill of the person performing the analysis (Heinicke et al., 2015). This may be a viable option with a skilled ecologist and a small dataset. However, the latter is becoming increasingly rare with advancing technology, and therefore the need for automated techniques is growing rapidly. Fortunately, automated techniques are experiencing notable improvement in terms of both accuracy and efficiency, largely due to the use of machine learning (Digby et al., 2013). Most automated tools utilise supervised machine learning and related methods, including artificial neural networks (Walters et al., 2012), random forest (Zamora-Gutierrez et al., 2016), Hidden Markov Models (Zilli et al., 2014), and support vector machines (Heinicke et al., 2015). These methods commonly use libraries of species calls or other sounds to facilitate detection when presented with new recordings. Currently, the low accuracy of these systems means that full automation is rare, and manual validation is often required (Kalan et al., 2016). However, new methods such as unsupervised feature extraction (Stowell and Plumbley, 2014) and deep convolutional neural networks (Goëau et al., 2016) can learn to classify directly from spectrogram data, often making them more robust and resistant to noise. At present, the main limitation for deep convolutional neural networks is large, clean datasets to train on.

1.6 Aims

1. To use data provided by Hill et al., 2018 to train a deep convolutional neural network that can detect gunshots in acoustic data
2. To investigate the effectiveness of using machine learning in cases such as this
3. To identify the presence of any spatio-temporal patterns of hunting on the Osa Peninsula

Methods

Sound is the propagation of waves of pressure through a medium. When a gun is fired, the vibrations produced alternately compress and rarefy the medium, leading to waves of high and low pressure that propagate in all directions (Bradbury and Vehrencamp, 2011). Over time and distance, these waves attenuate, their amplitude reducing as energy dissipates into the environment (Russ, 2013). The sound waves can then be transduced into an electrical signal. To record digitally, the analogue signal is sampled at a certain rate (typically measured in thousands of samples per second, kHz) and bit-depth (number of possible amplitude levels, typically 16-bit), with both parameters being important for determining frequency and amplitude resolution respectively. The signal information is then electronically recorded in the time-amplitude domain and can be processed mathematically using a fast Fourier transform (FFT) to convert the amplitude data into frequency data (Fourier, 1822; Cooley and Tukey, 1964). For a given time window in the recording, the FFT calculates the frequency components of the signal and their relative amplitudes, producing a frequency spectrum. For a visual representation of the whole recording, an FFT is calculated with an overlapping short sliding window across the length of the recording, producing a spectrogram (Society, 2003) This entire process is outlined in Figure 2.1.

It is common in audio classification to plot a variant of the traditional spectrogram: the mel-frequency spectrogram. This involves calculating the mel-frequency cepstrum (MFC) which is a representation of the sound's power spectrum after the frequency has passed through a mathematical function. Mel-frequency cepstral coefficients (MFCCs) are the constituent coefficients of an MFC (Xu et al., 2004). They are calculated from a cepstral representation of the sound, with frequency bands equally spaced on the mel scale (Stevens, Volkmann, and Newman, 1937). Spectrograms are fundamental to the analysis of acoustic data as they allow very specific sounds, such as a spider monkey call or a gunshot, to be visually identified and labelled, either manually or automatically.

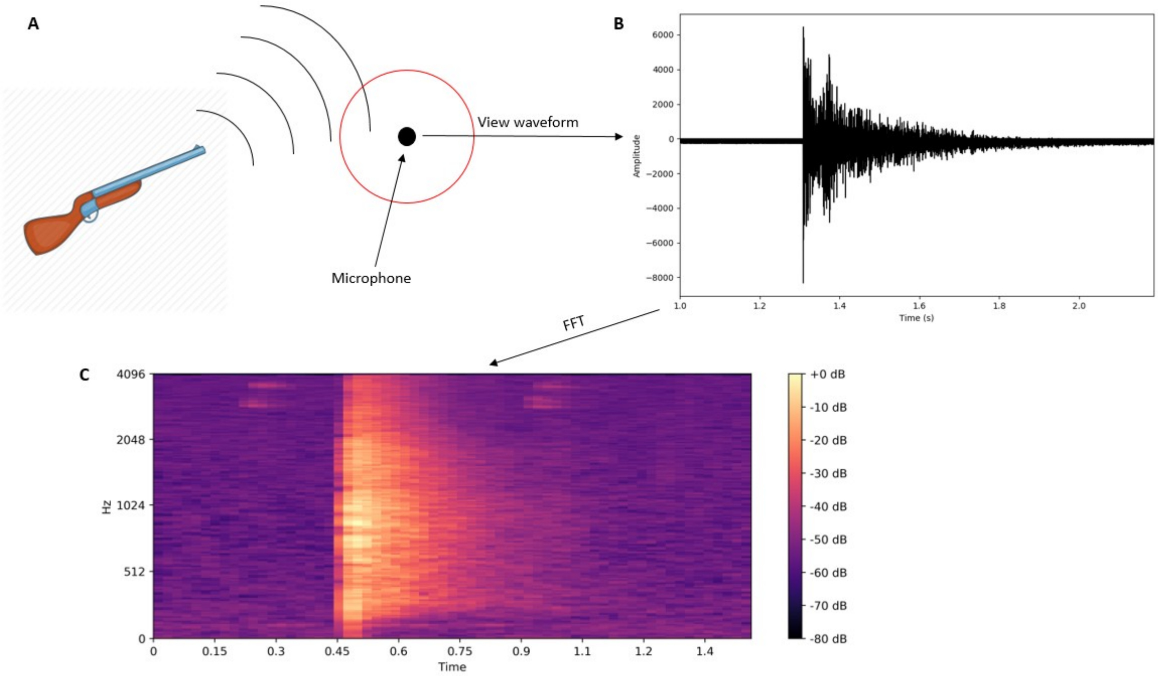

FIGURE 2.1: Recording and analysis of an acoustic signal. The emitted sound is transduced into an electrical signal by a microphone (**A**). In digital recording, the sound can be reconstructed in the time-amplitude domain (**B**) using the specified sampling rate (kHz). A frequency spectrum can then be produced using a fast Fourier transform (FFT), which calculates the signal's frequency components and their relative amplitudes. Calculating FFT within a sliding window across the recording produces a spectrogram, with time on the x-axis, frequency on the y-axis, and with amplitude (energy) shown as colour intensity (**C**)

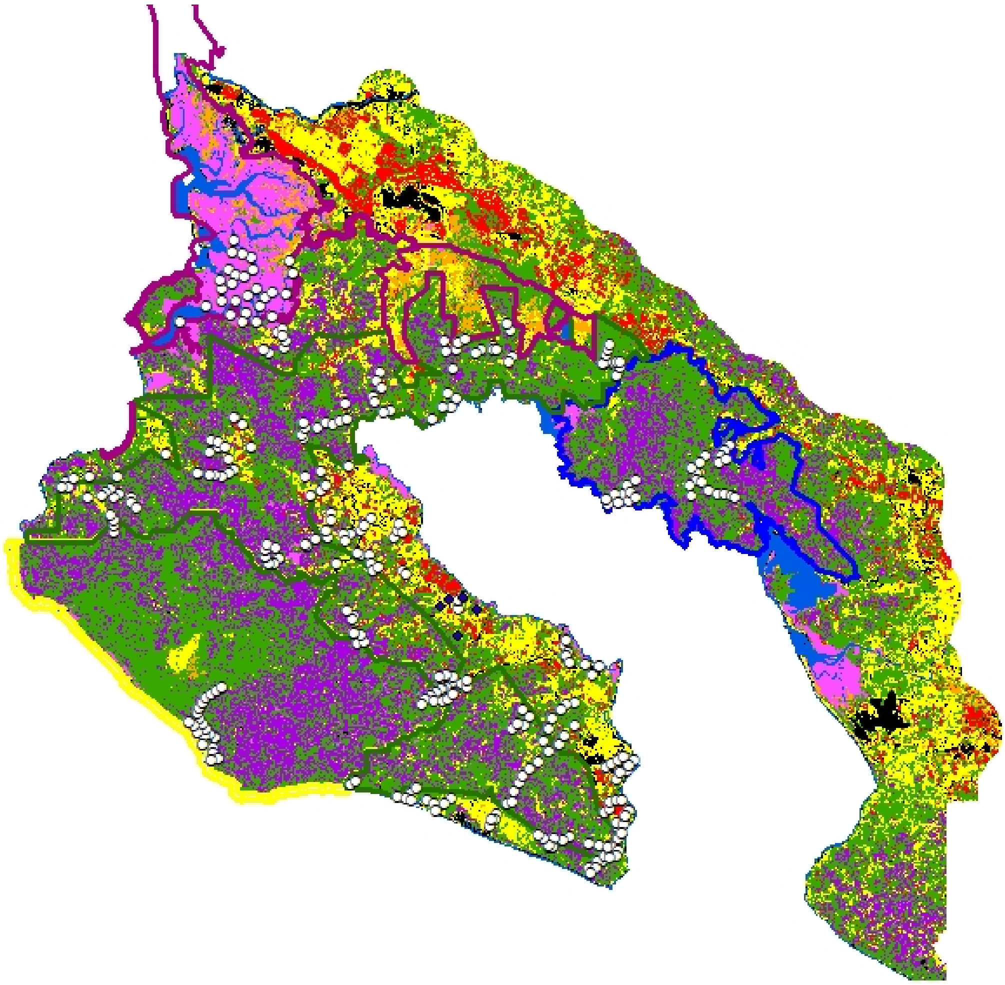

FIGURE 2.2: Map of the study site in the Osa Peninsula Lawson, 2019
(NEED KEY/AREAS MARKED)

2.1 Data collection and preprocessing

2.1.1 Study site

The audio data used in this study was collected at a study site in the Osa Peninsula, Costa Rica (Figure 2.2). This 2,500 km^2 site sits in a particularly diverse region, containing approximately 2.5% of the world's species on less than 0.001% of its land area (Lawson, 2019). Three national parks - Corcovado, Piedras Blancas, and the Terreba-Sierpe wetlands - and one forest reserve are represented within this study site. There has been significant landscape alterations in these areas which are inevitably brought about by an expanding human population and the introduction of agriculture and urbanisation (Höbinger et al., 2012).

2.1.2 Sampling

In this study, data from six AudioMoth acoustic sensors were collocted over six consecutive days, totalling 960 hours. These sensors covered protected forest reserve, protected grassland, and unprotected forest (TABLE), and were placed in areas known to be hotspots for hunting (personal communication, Jenna Lawson). Each device was set to record for three periods a day, 0500-0930, 1400-1830, and 2100-0300. These periods were chosen to coincide with the peak activity levels of *A. geoffroyi* and their associated poaching. Sound was recorded continuously during these periods at a sample rate of 48 kHz and the devices were contained in waterproof casing to avoid damage. Every minute of audio data was saved in a separate file with the filename being a Unix hexadecimal timestamp. Where possible, trails and other areas of high human density were avoided.

2.1.3 Training data

Hill et al., 2018 provided the labelled audio data which was used in their study and was used in this study to train the neural network. The data was collected using 36 AudioMoth devices in the tropical rainforests in Pook's Hill Reserve, Belize. The devices covered 13 sites and were placed 200 m apart from each other. Controlled gunshots were set off by the research team and were labelled as positives in the dataset through visual inspection of spectograms. The detection algorithms used were able to detect 66% of gunshots up to 500 m away and 50% up to 1 km away. Devices facing towards the gunshot were 80% more likely to detect it than those facing away.

2.2 Machine learning

All coding in this study was carried out using either Python (3.6.8) or bash (4.4.20), on an Ubuntu (18.04.3) operating system. For the training data, each audio file was split into four second clips using the pydub (0.23.1) Python module. Each clip was then plotted as a 60x60 mel-spectrogram using the librosa (0.7.0) Python module.

The CNN was then trained using the Keras (2.2.4) Python module. Mel-spectrogram images were imported and the pixel values were normalised to take values between 0 and 1 (rather than 0 and 255) as this has been shown to improve convergence and stability (Liao and Carneiro, 2016). The data was split randomly into training and validation data in the ratio 3:1 as this has been shown to be appropriate for a dataset of this size (Guyon, 1997). The seed for random number generation was set to ensure repeatability.

2.2.1 Convolutional Neural Network

The CNN was constructed using Chollet, 2016's tutorial as a template. The first layer was a convolutional 2D kernel which was convolved with the input layer to produce a tensor of outputs, the input layer being n spectrograms of 64x64x3 size. It contained 32 nodes which determines the dimensionality of the output space and a 3x3 convolutional window. A 'relu' activation function was then applied as this was shown by Glorot, Bordes, and Bengio, 2011 to enable better training of deep neural networks, compared to other common activation functions such as the logistic sigmoid. The identified features are then passed through a max pooling layer which determines the most activated presence of each feature within a 2x2 cluster of features. Essentially, this filters out the less important features and keeps the most important ones to be passed to the next layer. In this model, these first three layers were repeated in the same order two further times, with the model then totalling nine layers. The features were then passed to a 'flattening' layer. This is a layer that takes a two-dimensional matrix of features and transforms them into a vector of features that can be fed to a neural network classifier, in this case a 'dense' layer composed of 64 fully-connected neurons. These neurons linearly take all the inputs from the previous layer, apply a weight to them and output to the next layer which, in this CNN, was another 'relu' activation layer. The activation layer output was then passed through a 'dropout' layer. In this case, that involved randomly discarding 50% of nodes in an effort to minimise overfitting. The remaining nodes were put through another 'dense' layer, this time with 2 nodes as this CNN was only being trained in a binary 'gunshot' or 'no gunshot' manner. The output of this 'dense' layer was passed through a final 'softmax' activation layer which is similar to logistic regression but usually used in multi-classification problems. However, it has been shown to be more effective than logistic regression, even in binomial classification (FIGURE NEEDED).

2.2.2 Training and testing the model

Initially, I trained the model for 100 epochs on the data provided by Hill et al., 2018 by randomly splitting the data into training and test data in a ratio of 70:30. The training and test data was comprised of equal numbers of positive (gunshot present) and negative (gunshot not present) spectra as imbalanced ratios have been previously shown to be ineffective (Kim and Kim, 2018) and preliminary experimentation on this dataset confirmed this. This model was then fed the subset of data from the Osa Peninsula and I manually checked the returned 'gunshots' for authenticity. The validated gunshots were then used to retrain the model in the same manner, before being fed data from the Osa Peninsula that had not already been used to train the model. The model was retrained a further two times, once with a combined training dataset of both Hill et al., 2018 and data from the Osa Peninsula, and another with just the Osa Peninsula data, but this time the negatives used were the false positives originally identified. The number of returned ’gunshots’ by each of the four models is shown in TABLE. Further, I listend to 24 hours of acoustic data from one of the sensors in an attempt to identify any false negatives.

Results

3.1 Machine learning

The convolutional neural network (CNN) was initially trained on the labelled dataset provided by Hill et al., 2018 and this proved highly successful, with 95% accuracy acheived when presented with test data from the same dataset. I then fed data from the Osa Peninsula through this model and it returned 6912 'gunshots'. I manually validated these 'gunshots' through auditory inspection and determined that 252 were authentic gunshots, giving a model precision of 3.65%. I then re-trained the CNN on the new, manually-labelled dataset from the Osa Peninsula, and tested it by feeding through data from the Osa Peninsula that had not been used to train it. This model returned approximately five times as many 'gunshots'. I then assembled a combined training dataset, consisting of the data provided by Hill et al., 2018 and the new labelled data from the Osa Peninsula. The model trained on this dataset returned approximately 15 times more 'gunshots' then the original Belize model. Finally, I assembled a new training dataset consisting only of labelled data from the Osa Peninsula, but this time the spectra labelled as negatives were false positives that I had previously identified. Upon testing, this model returned about 4% fewer 'gunshots' than the original Belize model. The number of 'gunshots' found by each variant of the model is highlighted in Figure 3.2.

3.2 Spatio-temporal analysis

Using only the authenticated gunshots from the Belize model, I undertook further analysis to explore the temporal patterns of hunting (Figure **??**). Firstly, a chi-square test of goodness-of-fit was performed to determine whether gunshot frequency was independent of the day of the week. Gunshot frequency was equally distributed across the weekdays, X^2 (25, N = 252) = 42, p = 0.227. I then conducted a negative binomial regression on the effect of day-of-the-week on gunshot frequency as this has been shown to be an appropriate method for modelling overdispersed count data (Zeilis, Kleiber, and Jackman, 2008). This model showed that there is evidence that the days 'Monday' ($e^{estimate}$ = 2.50, z = 2.62, p < 0.009), 'Wednesday' ($e^{estimate}$ = 3.53, z = 2.85, p < 0.004), 'Friday' ($e^{estimate}$ = 3.47, z = 2.81, p < 0.005), and 'Saturday' ($e^{estimate}$

= 2.67, z = 2.81, p < 0.029) are significant predictors of gunshot frequency.

I then conducted similar analysis on the time of day gunshots occured (Figure **??**). Gunshots were assigned to either 'morning', 'afternoon', or 'night', mirroring the three periods of the day that the audio sensors were set to record. A chi-square test of goodness-of-fit was performed to determine whether gunshot frequency was independent of time of day. Gunshot frequency was equally distributed across the three periods of the day, X^2 (4, N = 252) = 6, p = 0.199. As before, I carried out negative binomial regression on the effect of time of day on gunshot frequency. This model showed that there is evidence that 'Morning' is a significant predictor of gunshot frequency ($e^{estimate}$ = 3.07, z = 7.07, p < 0.0001). 'Night' was also shown to be a significant predictor of gunshot frequency ($e^{estimate}$ = 0.26, z = -4.75, p < 0.0001).

Gunshot frequency between the six study areas was also compared (Figure **??**). A chi-square test of goodness-of-fit was performed to determine whether gunshot frequency was independent of area. Gunshot frequency was equally distributed across the six areas, X^2 (25, N = 252) = 30, p = 0.224.

	Estimate	Exponential	S.E.	z-value	95% C.I.	p-value
Monday	0.92	2.50	0.35	2.62	from 1.24 to 4.95	0.009
Tuesday	0.73	2.07	0.46	1.59	from 0.85 to 5.14	0.11
Wednesday	1.26	3.53	0.44	2.85	from 1.50 to 8.56	0.004
Thursday	0.79	2.20	0.46	1.73	from 0.91 to 5.45	0.083
Friday	1.24	3.47	0.44	2.81	from 1.47 to 8.41	0.005
Saturday	0.98	2.67	0.45	2.18	from 1.11 to 6.54	0.029
Sunday	0.38	1.47	0.47	0.81	from 0.58 to 3.74	0.42

TABLE 3.1: NegBin model. Estimated dispersion: 3.02 ($se = 1.09$).

	Estimate	exp.estimate	S.E.	z-value	95% C.I.	p-value
Morning	1.12	3.07	0.16	7.07	from 2.26 to 4.22	< 0.0001
Afternoon	-0.22	0.80	0.24	-0.95	from 0.50 to 1.27	0.34
Night	-1.34	0.26	0.28	-4.75	from 0.15 to 0.45	< 0.0001

TABLE 3.2: NegBin model. Estimated dispersion: 1.41 ($se = 0.35$).

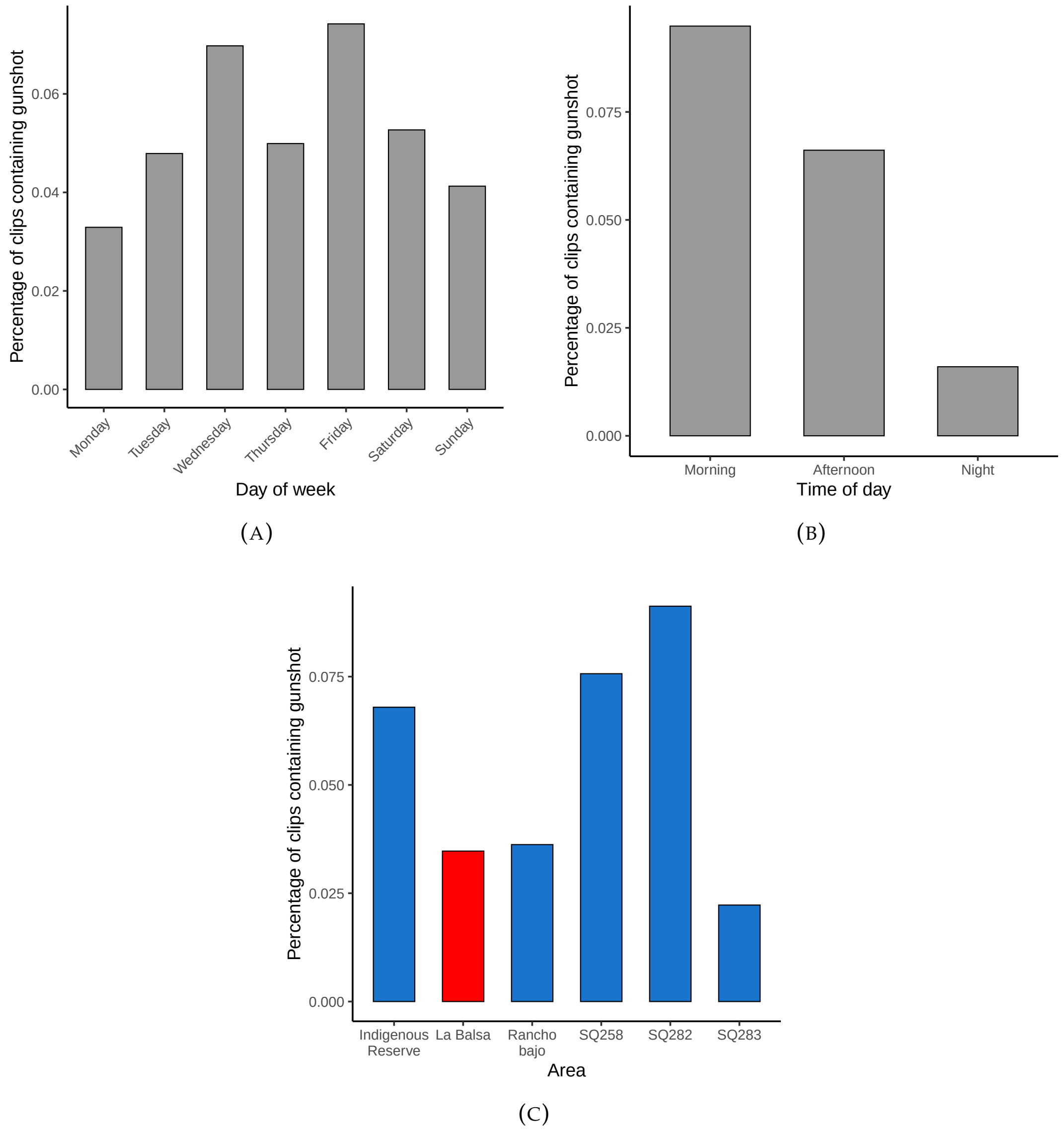
Percentage of clips containing gunshot
0.00
0.02
0.04
0.06
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of week
(A)
Percentage of clips containing gunshot
0.000
0.025
0.050
0.075
Morning
Afternoon
Night
Time of day
(B)
Percentage of clips containing gunshot
0.000
0.025
0.050
0.075
Indigenous Reserve
La Balsa
Rancho bajo
SQ258
SQ282
SQ283
Area
(C)

FIGURE 3.1

Area	Belize	Osa Peninsula	Combined	False Positives
SQ258	1179	2063	18080	1118
SQ283	1594	1343	12188	1576
SQ282	936	2404	22975	863
La_Balsa	758	5230	11108	729
Rancho_bajo	1514	3471	18003	1486
Indigenous_reserve	931	22565	27924	887

FIGURE 3.2: Number of returned 'gunshots' from each version of the convoluted neural network (CNN). The 'Belize' model was trained only on data provided by Hill et al., 2018, 'Osa Peninsula' was trained on data collected by Jenna Lawson (Lawson, 2019), 'Combined' was trained on a combination of the previous two datasets, and 'False Positives' was trained on Jenna's data but all negatives provided were previously identified false positives.

Discussion

4.0.1 Machine learning and its effectiveness

Training the CNN on the data provided by Hill et al., 2018 was very effective and allowed for very accurate classification of their data. Despite the high precision when presented with data from Belize, the model's precision took a significant hit when presented with data from the Osa Peninsula. This may be because, whilst the Belize acoustic data was also from a tropical forest in South America, it was from a different country with different fauna, weather, and soundscape. However, the model was still able to identify 246 gunshots which I manually checked for authenticity. This is a promising result, and with some optimisation may lead to greater precision and effectiveness for the use of neural networks and PAM going forward. Re-training the CNN on the manually-labelled data from the Osa Peninsula resulted in a significant increase in the number of 'gunshots' returned. This is likely to be indicative of lower precision; however, this cannot be stated as a certainty as the returned 'gunshots' were not manually validated. Re-training the CNN on a combined dataset of data from Belize and the Osa Peninsula resulted in even more 'gunshots' being identified. It is very likely that this is a result of increased numbers of false positives and in turn, lower precision, as this model classified approximately 25% of all four second clips as containing a gunshot, which seems doubtful. Again, manual validation would be required to support this assertion. Interestingly, the fourth dataset used to train the CNN - composed of authenticated positives and previously identified false postives labelled as negatives from the Osa Peninsula data - returned about 4% fewer 'gunshots'. This perhaps shows that highlighting to the CNN sounds that are similar to gunshots, such as a branch snapping, as being negative, improves the precision of the model. In this case, manual validation would also be required to support these claims. This strongly suggests that the primary factor in determining the effectiveness of a CNN model in studies such as this is the quality of the dataset used to train the model on. Going forward, the logical next step in this study area would be to increase the size of the dataset and to continue to retrain the model on increasingly large - and hopefully more accurate - data. This is likely the best way to improve the classification accuracy in the future.

4.0.2 Spatio-temporal hunting patterns on the Osa Peninsula

The temporal analyses carried out on the authenticated gunshot data returned some significant and interesting results. Firstly, certain weekdays proved to be significant indicators of gunshot frequency, which may be due to reasons similar to those suggested by Astaras et al., 2017, who highlighted the low frequency of gunshots on known market days. For Astaras et al., 2017, this implied that the majority of the hunting taking place was so that the meat could be sold for profit, rather than taking place for sport or sustenance, and this may also be the case for the hunting taking place on the Osa Peninsula. Regardless of the reasons behind the variation, this type of data will undoubtedly be useful for rangers patrolling these national parks if they are able to focus their efforts on certain days of the week.

Furthermore, time of day proved to be a very significant indicator of gunshot frequency, with over three times as many gunshots detected in the morning (0500 - 0930) compared to during the night (2100 - 0300). This is a stark contrast to the findings of Astaras et al., 2017 who found the complete opposite, with far more gunshots being detected during the night. As this study has used very similar techniques to the study undertaken by Astaras et al., 2017, there is presumably a difference between the two areas themselves causing the diverging trends. This difference may be patrolling frequency. Astaras et al., 2017 postulated that the increased hunting rate at night was due to regular ranger patrols during the day. Patrols in the national parks of the Osa Peninsula are said to be infrequent, and, when they do occur, only cover very small areas (personal communication, Jenna Lawson, 2019). This perhaps means that the patrols are an insufficient deterrent to hunting during the day, and the (probably low) risks of being caught during patrols are taken rather than hunting at night in forests that are treacherous and difficult to navigate. It should be noted, however, that despite much more regular and thorough patrolling in the study site of Astaras et al., 2017, the rangers' efforts do not seem to reduce the amount of illegal hunting occurring, but merely shift it to a different time of day, meaning that it takes place at night rather than during the day. Thus, the findings of this study ought to be considered in the management strategies of protected forests in the future, possibly by establishing night-time patrols in addition to day-time patrols.

Conclusion

The recent developments in passive acoustic monitoring (PAM) and machine learning techniques have opened a new door for conservation efforts around the world. This study has demonstrated the potential of these techniques for improving efficiency, both in terms of time and resources, aiding efforts which are constantly limited by these factors. The use of convolutional neural networks (CNNs) has enabled a partial automation of the detection of gunshots at this study site in Costa Rica, but an element of manual validation is still required. The accumulation of larger, more-extensive datasets is imperative moving forward, as this has been shown to be a key factor in determining the success of methods such as this.

Furthermore, in the data that was analysed in this study, clear patterns in the illegal hunting activity within this region were highlighted. In particular, time seems to be an important factor, with both the day of the week and time of day proving significant indicators of illegal hunting frequency. These findings in particular are significant, as they are able to immediately benefit the conservation of wildlife on the Osa Peninsula. They also provide a framework for similar studies in areas of conservation globally.

Bibliography

Allan, Eric et al. (2015). "Land use intensification alters ecosystem multifunctionality via loss of biodiversity and changes to functional composition". In: *Ecology Letters* 18.8, pp. 834–843. ISSN: 14610248. DOI: 10.1111/ele.12469.

Aquino, Rolando et al. (2013). "Distribution and abundance of white-fronted spider monkeys, Ateles belzebuth (atelidae), and threats to their survival in Peruvian Amazonia". In: *Folia Primatologica* 84.1, pp. 1–10. ISSN: 00155713. DOI: 10.1159/000345549.

Astaras, Christos et al. (2017). "Passive acoustic monitoring as a law enforcement tool for Afrotropical rainforests". In: *Frontiers in Ecology and the Environment* 15.5, pp. 233–234. ISSN: 15409309. DOI: 10.1002/fee.1495.

Barlow, Jos et al. (2016). "Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation". In: *Nature* 535.7610, pp. 144–147. ISSN: 14764687. DOI: 10.1038/nature18326. URL: http://dx.doi.org/10.1038/nature18326.

Bradbury, Jack and Sandra Vehrencamp (2011). *Principles of Animal Communication*. 2nd ed. OUP, USA.

Browning, Ella et al. (2017). *Passive acoustic monitoring in ecology and conservation*. Tech. rep., pp. 1–75.

Cardinale, Bradley J et al. (2012). "Biodiversity loss and its impact on humanity Bradley". In: *Nature* 486.4701, pp. 59–67. DOI: 10.1038/nature11148.Access.

Ceballos, Gerardo et al. (2015). "Accelerated modern human-induced species losses: Entering the sixth mass extinction". In: *Science Advances* 1.5, pp. 1–5. ISSN: <null>.

Chapman, A. D. (2009). *Numbers of Living Species in Australia and the World*. 2nd. Toowoomba, Australia: Australian Biological Resources Study, p. 3. ISBN: 978 0 642 56861 8.

Chapman, C. A., L. J. Chapman, and R. L. McLaughlin (1989). "Multiple central place foraging by spider monkeys: travel consequences of using many sleeping sites". In: *Oecologia* 79.4, pp. 506–511. ISSN: 1432-1939. DOI: 10.1007/BF00378668.

Chiarucci, Alessandro, Giovanni Bacaro, and Samuel M. Scheiner (2011). "Old and new challenges in using species diversity for assessing biodiversity". In: *Philosophical Transactions of the Royal Society B: Biological Sciences* 366.1576, pp. 2426–2437. ISSN: 09628436. DOI: 10.1098/rstb.2011.0065.

Chollet, Francois (2016). *Building powerful image classification models using very little data*. URL: https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html (visited on 08/18/2019).

Cooley, James W. and John W Tukey (1964). "An Algorithm for the Machine Calculation of Complex Fourier Series Author (s): James W . Cooley and John W . Tukey Published by : American Mathematical Society Stable URL : http://www.jstor.org/stable/2003354 REFERENCES Linked references are available o". In: *Mathematics of Computation* 19.90, pp. 297–301.

Cuarón, A. D. et al. (2008). *Ateles geoffroyi , Geoffroy's Spider Monkey*. URL: http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T2279A9387270.en{\%}0ACopyright: (visited on 07/22/2019).

De Vos, Jurriaan M. et al. (2015). "Estimating the normal background rate of species extinction". In: *Conservation Biology* 29.2, pp. 452–462. ISSN: 15231739. DOI: 10.1111/cobi.12380.

Digby, Andrew et al. (2013). "A practical comparison of manual and autonomous methods for acoustic monitoring". In: *Methods* 4, pp. 675–683. DOI: 10.1111/2041-210X.12060.

Doherty, Tim S. et al. (2016). "Invasive predators and global biodiversity loss". In: *Proceedings of the National Academy of Sciences* 113.40, pp. 11261–11265. ISSN: 0027-8424. DOI: 10.1073/pnas.1602480113.

Fanin, Nicolas et al. (2018). "Consistent effects of biodiversity loss on multifunctionality across contrasting ecosystems". In: *Nature Ecology and Evolution* 2.2, pp. 269–278. ISSN: 2397334X. DOI: 10.1038/s41559-017-0415-0. URL: http://dx.doi.org/10.1038/s41559-017-0415-0.

Fourier, J. B. J. (1822). *Théorie analytique de la chaleur*. Vol. 5. 121. Chez Firmin Didot, père et fils. DOI: 10.1038/005319a0.

Gibb, Rory et al. (2019). "Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring". In: *Methods in Ecology and Evolution* 10.2, pp. 169–185. ISSN: 2041210X. DOI: 10.1111/2041-210X.13101.

Glorot, Xavier, Antoine Bordes, and Yoshua Bengio (2011). "Deep Sparse Rectifier Neural Networks Xavier". In: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. Vol. 15, pp. 315–323. ISBN: 1532-4435. DOI: 10.1.1.208.6449. arXiv: 1502.03167. URL: http://proceedings.mlr.press/v15/glorot11a/glorot11a.pdf.

Goëau, Hervé et al. (2016). "LifeCLEF Bird Identification Task 2016: The arrival of Deep learning." In: *Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum*, pp. 440–449. URL: https://hal.archives-ouvertes.fr/hal-01373779/document.

Guyon, Isabelle (1997). *A scaling law for the validation-set training-set size ratio*. Tech. rep., pp. 1–11. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.33.1337{\&}rep=rep1{\&}type=pdf.

Heinicke, Stefanie et al. (2015). "Assessing the performance of a semi-automated acoustic monitoring system for primates". In: *Methods in Ecology and Evolution* 6, pp. 753–763. DOI: 10.1111/2041-210X.12384.

Hill, Andrew P. et al. (2018). "AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment". In: *Methods in Ecology and Evolution* 9.5, pp. 1–13. ISSN: 2041210X. DOI: 10.1111/2041-210X.12955.

Höbinger, Tamara et al. (2012). "Impact of oil palm plantations on the structure of the agroforestry mosaic of La Gamba, southern Costa Rica: Potential implications for biodiversity". In: *Agroforestry Systems* 85.3, pp. 367–381. ISSN: 01674366. DOI: 10.1007/s10457-011-9425-0.

Kalan, Ammie K et al. (2016). "Passive acoustic monitoring reveals group ranging and territory use : a case study of wild chimpanzees (Pan troglodytes)". In: *Frontiers in Zoology* 13.34, pp. 1–11. ISSN: 1742-9994. DOI: 10.1186/s12983-016-0167-8. URL: http://dx.doi.org/10.1186/s12983-016-0167-8.

Kim, Jinseok and Jenna Kim (2018). "The impact of imbalanced training data on machine learning for author name disambiguation". In: *Scientometrics* 117.1, pp. 511–526. ISSN: 15882861. DOI: 10.1007/s11192-018-2865-9.

Lawson, Jenna (2019). "Sounds of the Spider Monkey : Using acoustics to investigate biodiversity, and land use and threats to the Geoffroy's spider monkey (Ateles geoffroyi) in Costa Rica".

Liao, Z. and G. Carneiro (2016). "On the Importance of Normalisation Layers in Deep Learning with Piecewise Linear Activation Units". In: *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–8.

Merchant, Nathan D. et al. (2015). "Measuring acoustic habitats". In: *Methods in Ecology and Evolution* 6.3, pp. 257–265. ISSN: 2041210X. DOI: 10.1111/2041-210X.12330.

Newbold, Tim et al. (2015). "Global effects of land use on local terrestrial biodiversity." In: *Nature* 520.7545, pp. 45–50. ISSN: 1476-4687. DOI: 10.1038/nature14324. URL: http://www.ncbi.nlm.nih.gov/pubmed/25832402.

Pacheco, Luis F. and Javier A. Simonetti (2000). "Genetic structure of a mimosoid tree deprived of its seed disperser, spider monkey". In: *Conservation Biology* 14.6, pp. 1766–1775. ISSN: 08888892. DOI: 10.1046/j.1523-1739.2000.99182.x.

Rocchini, Duccio et al. (2018). "Measuring β-diversity by remote sensing: A challenge for biodiversity monitoring". In: *Methods in Ecology and Evolution* 9.8, pp. 1787–1798. ISSN: 2041210X. DOI: 10.1111/2041-210X.12941.

Roosmalen, M. G. M. van (1985). "Habitat preferences, diet, feeding strategy and social organization of the black spider monkey (Ateles paniscus paniscus Linnaeus 1758) in Surinam". In: *Acta Amazonica* 15.

Russ, Jon (2013). *British Bat Calls: A Guide to Species Identification*. 1st ed. Pelagic Publishing.

Society, The Royal (2003). *Measuring biodiversity for conservation*. Tech. rep., pp. 1–65.

Stevens, S. S., J. Volkmann, and E. B. Newman (1937). “A Scale for the Measurement of the Psychological Magnitude Pitch”. In: *Journal of the Acoustical Society of America* 8.3, pp. 185–190.

Stowell, Dan and Mark D. Plumbley (2014). “Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning”. In: *PeerJ* 2, pp. 1–31. DOI: 10.7717/peerj.488.

Walters, Charlotte L. et al. (2012). “A continental-scale tool for acoustic identification of European bats”. In: *Journal of Applied Ecology* 49.5, pp. 1064–1074. ISSN: 00218901. DOI: 10.1111/j.1365-2664.2012.02182.x.

Xu, Min et al. (2004). “HMM-based audio keyword generation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3333, pp. 566–574. ISSN: 03029743.

Zamora-Gutierrez, Veronica et al. (2016). “Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design”. In: *Methods in Ecology and Evolution* 7.9, pp. 1082–1091. ISSN: 2041210X. DOI: 10.1111/2041-210X.12556.

Zeilis, A., C. Kleiber, and S. Jackman (2008). “Regression Models for Count Data in R Achim”. In: *Journal of Statistical Software* 27.8, pp. 1–25.

Zilli, Davide et al. (2014). “A hidden Markov model-based acoustic cicada detector for crowdsourced smartphone biodiversity monitoring”. In: *Journal of Artificial Intelligence Research* 51, pp. 805–827. ISSN: 10769757.