ICIMOD

WORKING PAPER

Exploring the potential of deep learning for classifying camera trap data: A case study from Nepal

CORRESPONDING AUTHORS

Tanuja Shrestha (tanuja.shrestha@gmail.com), Mir A Matin (mir.matin@unu.ed)

AUTHORS

Tanuja Shrestha[1,2], Mir A Matin[1,3], Vishwas Chitale[1,4], Samuel Thomas[1]

Abstract

Data from camera trap networks provide crucial information on various important aspects of animal presence, movement, and behaviour. This can play an important role in wildlife studies and in prioritising wildlife conservation and management interventions. However, manual processing of the large volumes of images captured is time and resource intensive. The potential for automation of image recognition for wildlife through machine learning has not been adequately tested in the Hindu Kush Himalaya (HKH). In our study, we explore three different approaches of deep learning methods to detect and classify images of key animal species collected from the ICIMOD Knowledge Park at Godavari, Nepal. Our study demonstrates that transfer learning with ImageNet pretrained models (A1) can be used to detect animal species like porcupine, common leopard, and macaque with minimal model training and testing. Transfer learning as an integrated feature extractor (A2) appears beneficial for long-term wildlife monitoring efforts due to its ability to classify multiple animal species shown by our binary image classification tasks. Two data sets – 'Wild boar and Barking Deer' and 'Wild boar and Others' – were used, with which we achieved 95.71% and 83.24% accuracy, respectively. Likewise, our custom CNN model resembling VGG16 architecture (A3) showed 87% and 71% accuracy for the above datasets, which shows the potential to perform well when data size is increased. These methods when scaled up offer tremendous scope for quicker and informed conflict management actions, including automated response, which can help minimise human wildlife conflict management costs across countries in the region. They can also be extended to monitor animal habitat and behaviour. Automatic detection of animals could also enable mechanical solutions for repelling animals and reduce labour costs related to guarding and response.

Keywords: Machine learning, Artificial Intelligence, Convolutional Neural Network, Deep learning, Camera Traps, Image Classification, Object Detection

[1] International Centre for Integrated Mountain Development (ICIMOD); [2] Fortescue Metals Group, Australia; [3] Institute for Water, Environment and Health (UNU-INWEH); [4] Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH, India

Acknowledgements

We would like to thank our former Director General David Molden for urging us to start this survey, and Laurie Ann Vasily for bringing the trail cameras from the USA. We want to express our sincere gratitude to Pema Gyamtsho, our present Director General for his support. We also want to express our gratitude to the staff at the ICIMOD Knowledge Park, namely Samden Sherpa, Jiwan Tamang, Purna Bhujel, Santa Bahadur Tamang, and Basanta Tamang.

Contents

SECTION I

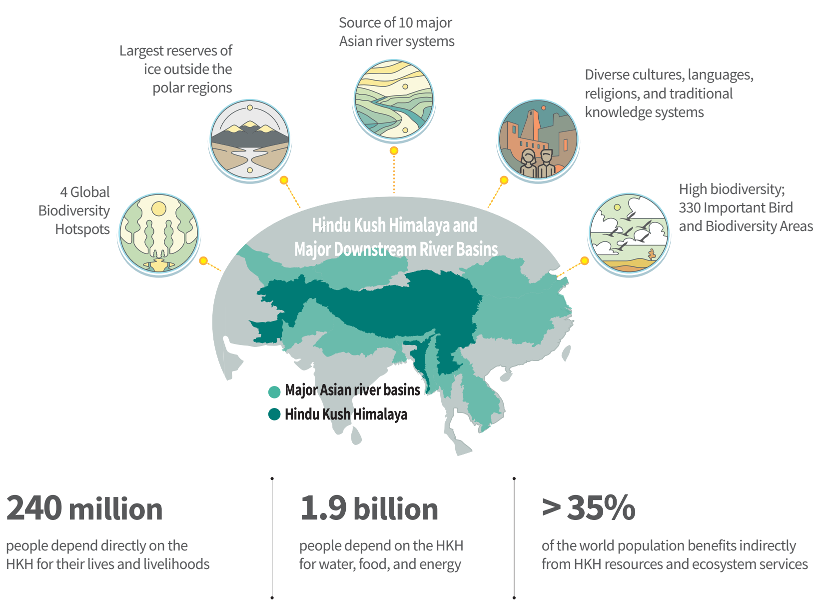
Introduction

1.1 Importance of deep learning/ machine learning in processing camera traps data in the region

The Hindu Kush Himalaya (HKH) comprises all or parts of eight South Asian countries, namely Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal and Pakistan. The region with an area of around 4.2 million km2 exhibits great heterogeneity in terms of climate, biodiversity, and topography and is home to diverse cultures, languages, religions and traditional knowledge systems. It has the largest reserves of snow and ice outside of the Polar Regions and contains the headwaters of ten major Asian river systems. The HKH includes four global biodiversity hotspots and over 330 important bird and biodiversity areas (IBAs). The HKH supports the livelihoods of some 1.9 billion people – 240 million living in the hills and mountains, and 1.65 billion in the river basins downstream – who depend on it for water, food, and energy, among other things. More than 35% of the global population benefits indirectly from its resources and ecosystem services (Figure 1).

Nearly 39% of the HKH total area is under the protected area (PA) network. However, many of these PAs are small, isolated, and lack corridor connectivity with each other. In mixed use landscapes, this creates conditions for human-wildlife interactions and conflict. The Kanchenjunga Landscape, a part of the Himalaya biodiversity hotspot and spread across parts of Nepal, Bhutan and India, is greatly affected by human-wildlife conflict, resulting in human and animal injury and death, and crop and livestock depredation by wildlife. Mammals such as black bear, civet, common leopard, snow leopard, dhole, tiger, golden jackal, yellow-throated marten, wolf, and eagles are reported to predate on livestock and poultry. Likewise, barking deer, elephant, porcupine, hare, sambar deer, sloth bear, wild pig, black bear, civet, jungle rat, macaque, squirrels and various birds are known to depredate crops. A large number of human casualties also result each year from conflict with elephants and this has led to retaliatory killing of elephants (Chettri and Gurung 2017).

FIGURE 1 NATURAL RESOURCES IN THE HKH REGION



Source: Sharma et al. (2019)

Narrowing down to country specific trends, Bhutan is experiencing a steep rise in HWC over the past decades, with livestock predation and crop depredation being major concerns. The country loses a huge amount of crops each year due to conflict species like wild pig, porcupine and macaque, causing significant economic losses to farming households (The Bhutanese 2013; Wangyel et al. 2019) Bhutan to assess the consequence of crop raided by conflict species in four villages located in the buffer zone. Data were collected through semi-structured household interviews, direct observation, and key informant interviews. The majority of households interviewed (95.5%. Other species reported to raid crops are barking deer, rhesus macaque (Macaca *mulatta*), and sambar (*Rus unicolor*). About a year ago, *Kuensel*, Bhutan's national daily reported that the government was establishing a HWC endowment fund of Nu16 million in order to help conflict-affected farmers beginning in 2020. The ministry is seeking support from the European Union and has agreed to contribute Nu16 million annually to the fund after it is established. The nation has set a target to collect at least USD 35 million and claims to have already collected Nu 40 million (KuenselOnline 2019).

In India, 90% of the country's geographical area has reported a rise in HWC. A study carried out in 2016 reported elephants, tigers and leopards causing damage to humans and other wildlife species (Anand and Radhakrishna 2017). Due to extensive crop damage by rhesus macaque, blue bull and wild boar, they have been declared vermin in the states of Himachal Pradesh (2016), Bihar (2015) and Uttarakhand (2016), which means that these species are not under any protection and can be killed under conflict circumstances (Anand and Radhakrishna 2017).

In the Kanchenjunga Conservation Area of Nepal, livestock depredation is a severe problem in Ghunsa Valley. The losses in Ghunsa from 2005 to 2014 stand at an annual rate of 11%. In 2017, the loss rate stood at 28%, an increase by approximately 11% since 2014 (17.2%), despite community-based insurance schemes. In Yamphudin, crop damage is a severe problem. Wild dogs are reported to cause much of the livestock depredation, whereas crop damage is caused by black bear, palm civet, barking deer, rhesus macaque and porcupine (Sherchan and Bhandari 2017)we review the extent to which wildlife damages livestock and crops.
 Location Lelep and Yamphudin region, Kanchenjunga Conservation Area, Taplejung, Nepal
 Materials and methodsThe study employed a combination of surveying methods such as focus group discussion, key informant interview and field observation from 21 July to 06 August 2013. Focus group discussion was done primarily with the representatives of snow leopard conservation committee in Lelep and Yamphudin.
 Key Findings Livestock depredation in Ghunsa valley, Lelep village development committee was increasing with an annual average loss rate of 11% in ten years (2005- 2014.

Therefore, the presence, behaviour and movement of these species should be researched on a large scale and in an automated manner to develop holistic strategies to manage various conflicts in the region.

1.2 Camera traps data in ecological studies

Images acquired from camera traps have become progressively important and inevitable in ecological management studies due to their potential for close monitoring of animal movement and behaviour. They are efficient in capturing wildlife images continuously and inexpensively. In particular, they are useful in remote observation of wildlife given the type of information captured in these images, such as the spatial location, time and date of the location of various animal species. These bits of information are instrumental in ecological analyses, like counting, behaviour monitoring, measuring population distribution and density, identifying vulnerabilities if any, habitat and niche mapping, overexploitation of natural resources, and human-wildlife conflict, among others (Peterson et al. 2019; Tabak et al. 2019).

The capture of images by trail cameras at certain intervals is triggered by the activation of an infrared sensor when movement is detected. As a result, the amount of data collected through the process is often large, sometimes reaching millions of images. For example, in Serengeti National Park, Tanzania, the Snapshot Serengeti Project deployed hundreds of camera traps to understand the dynamics of Africa's animal species. It is reported that from 2010 to 2013, the project collected 3.2 million images from 225 camera traps. The amount of human effort required to process, analyse and study such large amounts of data is considerable and therefore challenging to manage, despite the many benefits from camera traps (Norouzzadeh et al. 2018)detailed, and up-to-date information about the location and behavior of animals in the wild would improve our ability to study and conserve ecosystems. We investigate the ability to automatically, accurately, and inexpensively collect such data, which could help catalyse the transformation of many fields of ecology, wildlife biology, zoology, conservation biology, and animal behavior into "big data" sciences. Motion-sensor "camera traps" enable collecting wildlife pictures inexpensively, unobtrusively, and frequently. However, extracting information from these pictures remains an expensive, time-consuming, manual task. We demonstrate that such information can be automatically extracted by deep learning, a cutting-edge type of artificial intelligence. We train deep convolutional neural networks to identify, count, and describe the behaviors of 48 species in the 3.2 million-image Snapshot Serengeti dataset. Our deep neural networks automatically identify animals with >93.8% accuracy, and we expect that number to improve rapidly in years to come. More importantly, if our system classifies only images it is confident about, our system can automate animal identification for 99.3% of the data while still performing at the same 96.6% accuracy as that of crowdsourced teams of human volunteers, saving >8.4 y (i.e., >17,000 h at 40 h/wk.

1.3 Machine learning in ecological research

A significant number of current research projects have started using machine learning approaches, particularly deep learning, for extracting information from camera trap data, although the application of traditional ML and a few deep learning methods in wildlife monitoring studies date back to early 2000. Literature shows that ML approaches were pivotal in solving many complex ecological issues given their ability to process complex nonlinear data. Some of the studies included automatic classification of amphibian and bird calls, ecological modelling, and animal behaviour monitoring (Acevedo et al. 2009; Recknagel 2001; Valletta et al. 2017)decision tree, and support vector machines. These studies used machine learning algorithms, like support vector machine, random forest, decision trees, and artificial neural network. On the other hand, the current popularity and rise of deep learning technologies can be attributed to the modern high-tech advances attained in collecting large data sets and the subsequent demand for complementary and powerful data modelling methods (Valletta et al. 2017).

Among various current deep learning methods, Convolutional Neural Networks (CNN) have proven to be a powerful method in image classification, natural language processing, and speech recognition tasks. The two dimensional filters present in CNN layers can extract general and complex features with only little or no preprocessing of input data. In contrast, traditional machine learning approaches, such as support vector machine and random forest, required manual construction of hand-engineered features, making this a challenging task. Much of the success of deep learning methods can also be attributed to GPU (Graphics Processing Unit), massively data-parallel vector computations, and high-level open-source libraries, such as Keras, built on frameworks such as Tensorflow (Wu 2017).

To illustrate classification of camera trap data with deep learning, 48 animal species from 3.2 million images in the Serengeti Project were classified using convolutional neural network with a significant accuracy. This task, if done manually, would have required thousands of human volunteers. It is estimated that more than 8.4 years of manual effort was saved through automation (Norouzzadeh et al. 2018)detailed, and up-to-date information about the location and behaviour of animals in the wild would improve our ability to study and conserve ecosystems. We investigate the ability to automatically, accurately, and inexpensively collect such data, which could help catalyse the transformation of many fields of ecology, wildlife biology, zoology, conservation biology, and animal behavior into “big data” sciences. Motion-sensor “camera traps” enable collecting wildlife pictures inexpensively, unobtrusively, and frequently. However, extracting information from these pictures remains an expensive, time-consuming, manual task. We demonstrate that such information can be automatically extracted by deep learning, a cutting-edge type of artificial intelligence. We train deep convolutional neural networks to identify, count, and describe the behaviors of 48 species in the 3.2 million-image Snapshot Serengeti dataset. Our deep neural networks automatically identify animals with >93.8% accuracy, and we expect that number to improve rapidly in years to come. More importantly, if our system classifies only images it is confident about, our system can automate animal identification for 99.3% of the data while still performing at the same 96.6% accuracy as that of crowdsourced teams of human volunteers, saving >8.4 y (i.e., >17,000 h at 40 h/wk.

Similarly, Singh and Carpio et al. (2018) implemented deep learning methods to address animal-human cohabitation problems. In the study, researchers used pre-trained CNN model to classify animal species at edge devices. Similarly, long short-term memory neural network is used to predict the time-varying data traffic that aided in allocating bandwidth in fiber-wireless (FiWi) access networks. For passing vehicles, they generated an alert signal using wireless sensor networks where animals were frequently crossing roads.

Some studies have used deep learning methods to track illegal wildlife trade and poachers. A group of researchers from Finland and South Africa have discussed a framework that used deep learning methods to track illegal wildlife trade by filtering out information irrelevant to trade from user-generated content, such as images, text and videos. These data are publicly downloadable from various social media platform like Facebook, Twitter, and Flickr among others through application program interface and can be used for automated content classification. For instance, if the model finds the content as 'pangolin armored vehicle', then it could imply a sign of possible illegal trade. But the model discards phrases like 'pangolin taxa' (Di Minin et al. 2019).

The Teamcore Lab at the University of Southern California's Center for Artificial Intelligence in Society tested two approaches of artificial intelligence to beat poachers. The first one involves detecting poachers at night automatically in near real-time with an application called Systematic Poacher Detector using long wave thermal infrared cameras on drones and reporting to rangers through an alarm. The second one is called PAWS, Protection Assistant for Wildlife Security, a machine learning algorithm that predicts the places where poachers might set their traps. This is carried out by analysing past data on traps so that rangers can clear the traps before animals are harmed. PAWS was tested at 24 ranger patrolling areas in Srepok Wildlife Sanctuary, Cambodia. More than 10000 snares were discovered from mid-December 2018 to late January 2019. The figure was double the number before introduction of the AI software (Idea Connection 2019).

1.4 Significant amount of camera trap data in the HKH

There is a considerable amount of camera trap data that can aid in large scale automated animal recognition studies in the region. These data, however, have been generally used for studying key flagship and endangered species, forming a basis for further field visits and extended research on those particular species. This tendency has been reported in a study in India, with 72% of HWC studies focusing on only one taxonomic group, although a total of 88 species were reported to have been involved in conflict (Anand and Radhakrishna 2017). Also, camera trap data are often manually checked for the occurrence of one/several animal species in most studies; they are seldom analysed and processed remotely in an automated manner. This process has indirectly disregarded the true potential of deep learning that can aid in more comprehensive analysis of many species simultaneously. This could be attributed to the management challenge posed by large amounts of data created by each study. Another reason could be the limited knowledge on ML and deep learning studies within the region as compared to other regions.

SECTION II

Objectives

Therefore, given the availability of camera trap data, the severity of human-wildlife conflict, and the importance of advancing the use of machine learning/deep learning in the region, we conducted this pilot case study to test the potential for automatic identification of local wildlife species from camera trap data captured at ICIMOD Knowledge Park in Godavari, Nepal. Although our specific objective is to explore the ability of modern deep learning techniques in identification of animal species found in our study site, and reported to cause HWC in the region, we believe this study will play a fundamental role in supporting two future conservation goals in the HKH region. Firstly, it will shed light on the importance and usefulness of deep learning methods in wildlife surveillance, since the analysis and management of camera trap data using deep learning is relatively unexplored. Secondly, given the region's biodiversity richness, with many endemic and endangered animal species and ample amounts of camera trap data, the study could facilitate and encourage open access camera trap data archive, which is the present obstacle to conducting large scale monitoring efforts using machine learning. Several organisations that are direct or indirect partners of ICIMOD own large amounts of camera trap data and there is potential to create and promote an open access data archive for the HKH.

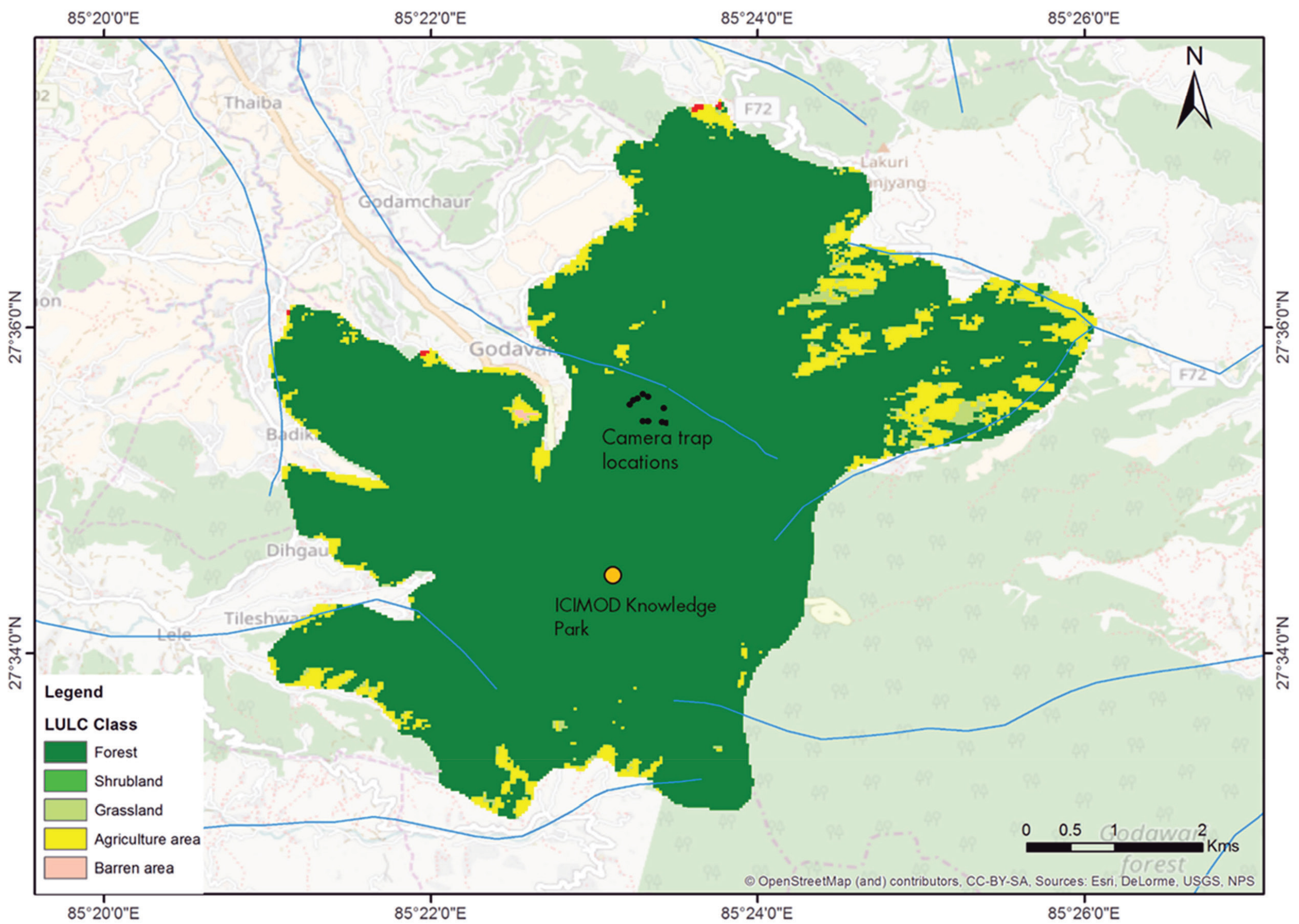
SECTION III

Methods

3.1 Study area

The ICIMOD Knowledge Park at Godavari was established in 1993 as a test and demonstration centre for sustainable mountain farming practices and appropriate technologies (Figure 2). It lies 15 km south of Kathmandu, Nepal, in the Phulchoki watershed area and is spread over approximately 30 hectares. The Godavari Kunda Community Forest lies to its northeast and the Diyale Community Forest to its southwest. The elevation of the Park ranges from 1,510 to 1,780 masl and slope from 0 to 60°C. The areas experiences subtropical and warm-temperate climate, with mean annual temperature of 17.2°C. It receives much of its

FIGURE 2 STUDY AREA



precipitation during the monsoon, with an annual average of 2000 mm. The vegetation is dominated by mixed deciduous and evergreen broadleaf species that have regenerated naturally. The steep slopes have natural forests, mixed slopes have shrubs, and valley floor has shrubs and bushes (Karki and Udas 2016).

3.2 Data capture

Our study used data from six trail cameras (Browning Trail Camera – Strike Force HD Pro). Two of them were deployed on 24 January 2019 and four on 1 February 2019 with the purpose of documenting the presence of wildlife. Wild pig (*Sus scrofa*), barking deer (*Muntiacus muntjak*), Himalayan or masked palm civet (*Paguma larvata*), large Indian civet (*Viverra zibetha*), yellow-throated marten (*Martes flavigula*), rhesus macaque (*Macaca mulatta*), Himalayan crestless porcupine (*Hystrix brachyura*), black-naped hare (*Lepus nigricollis*), leopard cat (*Prionailurus bengalensis*), jungle cat (*Felis chaus*), and common leopard (*Panthera pardus*) were captured.

The sampling was opportunistic and was mainly carried out to establish the presence of wildlife in the Park and to validate anecdotal evidence and reported sightings of various species by staff and visitors. The camera trap locations included wetland/stream, forest trails, human used trails and recognised animal paths – all chosen to maximise the likelihood of capture and establish presence/absence. The camera traps were moved around when there appeared to be signs of "camera shyness" and also to cover areas that had not been covered. Colour images were collected during the day without flash (4608*2560 resolution). At night, images were collected in grayscale settings (5760*3200 resolution). Videos were also collected to document animal behaviour.

3.3 Model architecture

3.3.1 VGG16

VGG16 is one of the top performing image classification models out of several image classification models available in the Keras API and a relatively simple model to implement. The model was created by the Visual Geometry Group at the University of Oxford and therefore named after them as "VGG16". It secured second position in the 2014 ImageNet Large-Scale Visual Recognition Challenge (ILSVRC). '16' denotes the number of weight layers in the model (Russakovsky et al. 2015).

ImageNet data comprises 1.2 million training, 50000 validating and 10000 testing images. Previously, the data was collected as part of the project in computer vision research to identify 22000 object categories. Later on, the project's objective was changed to identify 1000 classes and was submitted as a challenge publicly. The object categories include things that we use and see in our everyday lives, like various animals, household materials, vehicles, and many more. Most of the deep learning models progressed through running experiments on this data and achieved remarkable results leading to establishment of the top image classification models, one of them being VGG16. It secured 92.5% as its top 5 accuracy and 96% as top 1 accuracy in the 2014 challenge (Deng et al. 2009; Russakovsky et al. 2015).

VGG16 model has 13 convolutional layers spread across five blocks, 5 maximum pooling layers distributed within these five blocks, and 3 dense layers at the end, making a total of 21 layers. Number of total weight layers is 16 since pooling layers do not have weights. It has 64 filters in the first block, 128 in the second, 256 in the third, and 512 in the remaining ones. The filters' size is relatively smaller than the other image classification models with the window size of 3*3. The size in the volume is handled further down with the maximum pooling layers tucked in between various blocks. Dense layers consist of two fully connected layers with 4096 nodes and a last Softmax layer of 1000 nodes (Neurohive 2018).

3.3.2 ResNet50

ResNet stands for Residual Networks and is used for many computer vision tasks, speech recognition and natural language processing tasks. The model secured first position in the 2015 ImageNet and COCO competitions - ImageNet detection, ImageNet localisation, COCO detection, and COCO segmentation (Russakovsky et al. 2015). A special feature of this model is its ability to train very deep neural networks with around 150+ layers due to the presence of residual layers. The model has five stages with convolution and identity blocks. Each convolution and identity block has three convolution layers. Approximately over 23 million trainable parameters are present. The original input to the model is added to the output of the CONV, RELU, and BN layers. This process is referred to as identity mapping since the input (identity) is added to the output to the series of operations. The learning process can be mathematically symbolised as: $f(x) + id(x) = f(x) + x$, $id(x)$ is the identity function where the residual layer approximates y through $f(x)$, yet in traditional neural networks the learning function in general is $y = f(x)$ (Karim 2019).

3.4 Machine learning approaches

We explored three different machine learning approaches to classify selected animal species found in the Godavari site: i) Transfer learning with ImageNet pretrained weights (A1); ii) Transfer learning as an Integrated feature extractor (A2); and, iii) Custom CNN model resembling VGG16 architecture (A3). We used both VGG16 and ResNet50 model's pretrained weights in A1, whereas we used only VGG16 in A2. In A3, we customised the VGG16 model (Table 1).

TABLE 1 SUMMARY OF APPROACHES USED

Appro-aches	Architecture Used	Pretrained weights	Species
A1	VGG16 and ResNet50	Used	All species
A2	VGG16 (customised)	Used	Wild Boar and Barking Deer; Wild Boar and Others
A3	VGG16 (customised)	Not used	Wild Boar and Barking Deer; Wild Boar and Others

We tried implementing A1 in all species recorded. With A1 and A2, we conducted experiments with two data categories – 'Wild Boar and Deer' and 'Wild Boar and Others'. The number of images captured for the remaining species, except *Sus scrofa* (wild boar) and *Muntiacus muntjak* (barking deer, Indian muntjak), were significantly low due to which we have postponed the task of their recognition. We used the open source machine learning framework, Tensorflow, and neural network library, Keras, for the task. The sections that follow briefly describe these approaches and the results.

3.4.1 Transfer learning with ImageNet pre-trained weights (A1)

Transfer learning, carried out with pre-trained models, entails using knowledge achieved in one problem to a similar new problem, for instance, using a model trained to classify a car in order to recognise trucks. It is a growing field in many computer vision and natural language processing problems and has huge commercial potential. The fact that features extracted by the initial convolutional layers in the pre-trained models, trained on millions of images, are purely descriptive and general is the basis of this method's usability in deep learning research today. In addition, training deep learning models from scratch requires a huge amount of data, from a minimum of thousands to millions of images, given that is has to be trained to learn complex and challenging features and avoid overfitting of the model.. This has resulted in a significant data generation hurdle often requiring tedious data cleaning and massive computational resources to many scientists and researchers. Transfer learning addresses this problem by offering extensive and relevant expertise to related data sets (Weiss, Khoshgoftaar, and Wang 2016).

Keras Applications provide several top-performing pre-trained deep learning models with pre-trained weights. They can be readily downloaded and implemented for direct prediction, feature extraction and fine-tuning to a new task. Instantiating a model automatically downloads and stores the weights locally.

In this approach (A1), we used images of all animal species since it does not require model training and testing. ImageNet data comprises images of many animals similar to those found in the Godavari and Kanchenjunga Landscape that formed the basis for implementing this approach. Also, literature shows that the majority of deep learning research efforts in wildlife surveillance have implemented either the second or the third approach. Therefore, we explored this method and assessed its potential for classifying animal species found in the study area. This method gives labels from the ImageNet data set and provides output on any 'similar looking animals' within the ImageNet data set. Based on this, the top five predictions from the models were observed.

3.4.2 Transfer learning as an integrated feature extractor (A2)

Two categories of data sets, 'Wild Boar and Deer', and 'Wild Boar and Others' – of 11 mammals captured on the trail cameras, were used in this approach. Unlike A1 where training images are not required, A2 and A3 demanded preparation of quality training and testing images. Thus, the data were cleaned and videos were changed to images for classification purposes.

Preparing a custom dataset for wildlife pictures is a challenging process. Although animal species were captured by the videos, the animals' position was similar in most frames. Hence, images collected from the videos look similar (Figure 3). The more the variation in images, the more useful they are in building deep learning models. Many images were out of focus and taken from some distance, so we avoided using those images for training. This significantly affected the sample size despite the large collection of images. Some images and video frames did not capture a single animal species and may have been triggered by the movement of leaves, rain drops or insects flying across the sensor.

FIGURE 3 VIDEO FOOTAGE CONVERTED TO IMAGES (SIMILAR PHOTOS)

Therefore, as a solution, Tensorflow's Object Detection API with a pre-trained model was used to increase the number of images. The model was run with images where species were out of focus and noise was prominent. The bounding boxes acquired from the pre-trained model were clipped skipping the labels and scores and more zoomed images – different to original images - were prepared in an automated manner (Table 2). In this way, we avoided images where animals were out of focus, decreasing noise and eventually acquiring more images for training. 644 images were created for the 'Wild Boar and Deer' dataset from the first and second round of images and videos collected. For the 'Wild boar and Others' dataset, 300 images were used. The 'Other' category comprised species other than wild boar. 80% of images were kept for training, and 20% for testing for the first data set; for the second dataset, the ratio was 70% and 30%.

3.4.2.1 MODEL CONFIGURATION

We used the weights from the ImageNet data trained with VGG16 model and froze all the weight layers as non-trainable. This is referred to as the feature extracting process, where the features learned during training of the pre-trained model with ImageNet data are directed to the classification purpose. To do so, from the VGG16 model, we removed the fully connected layers from the output layer, and added a new fully connected layer and an output layer with 128 nodes and 1 node, respectively. Instead of three dense layers in a typical VGG16 model, our model has two dense layers followed by an output layer (Figure 4).

FIGURE 4 MODEL CONFIGURATION FOR A2 APPROACH

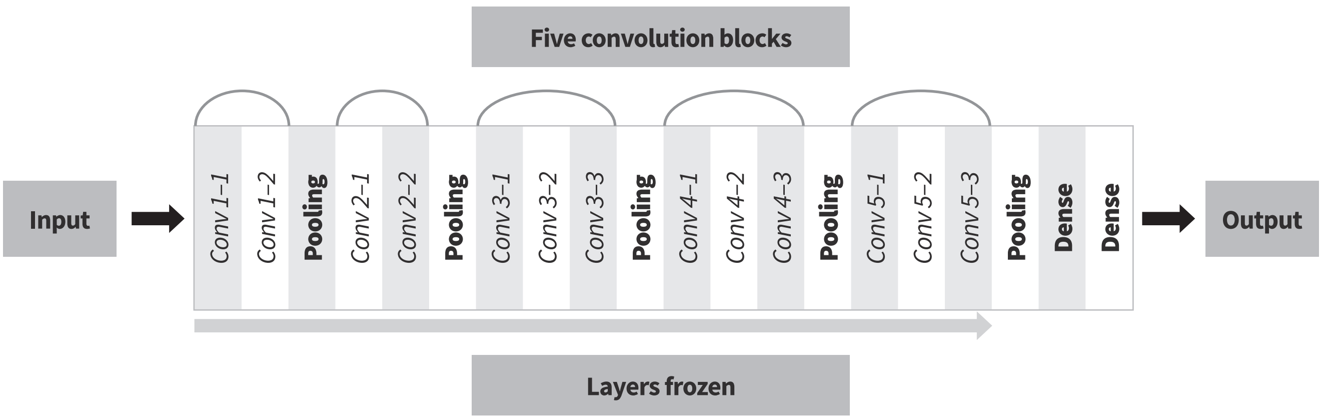
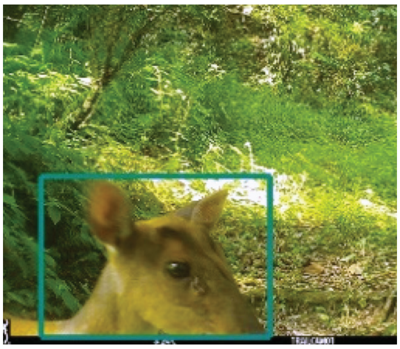


TABLE 2 DATA CLEANING WITH OBJECT DETECTION API

Original images with detection boxes	Bounding box clipped
	
Wild boar	
	
Wild boar	
	
Barking deer	
	
Barking deer	

3.4.2.2 IMAGE PREPROCESSING

Transfer learning as a feature extractor is not data intensive as opposed to training a model from the scratch, since all the weight layers use readily available weights from the pre-trained model. This also makes the data augmentation process easier and light weighted. The major augmentation measure implemented is feature wise centering to images. It involves subtraction of mean pixel values from every channels - red, green and blue – from the ImageNet data set. The problem type was specified as 'binary' and batch size was fixed to 64. The model was compiled with the stochastic gradient decent optimiser at the learning rate 0.001, momentum 0.9, and loss 'binary_crossentropy'. Number of iteration was set to 10.

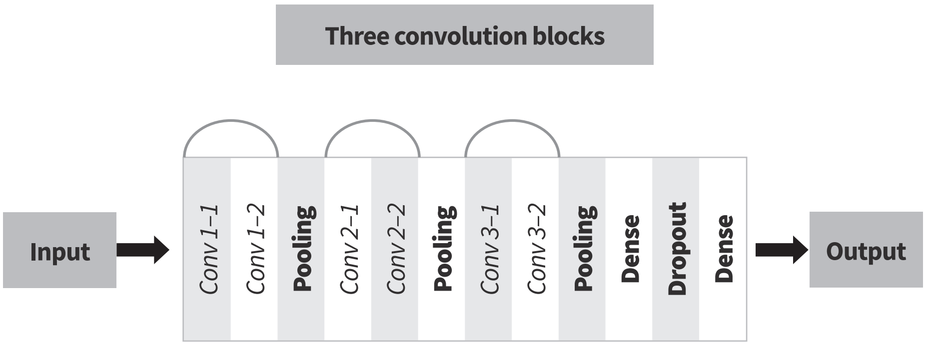
3.4.3 Custom CNN model resembling VGG16 architecture (A3)

In the third approach, we customised the model parameters but kept our model's architecture close to the VGG16. Although a high level of accuracy was achieved from our previous experiment with 95.71% accuracy for the 'Wild Boar and Deer' dataset, and 83.24% in the 'Wild Boar and Others' dataset, this method was carried out to observe the effect of the number of parameters or complexity offered by a deep model like VGG16 on model training and overall results. Another reason for adopting this approach was to understand the computational costs of running a model from scratch against transfer learning. We also wanted to compare the accuracy of this approach with transfer learning and come up with a model decision for the selected tasks.

3.4.3.1 MODEL CONFIGURATION

Instead of five convolutional blocks, three convolutional blocks of 3*3 sized filters with ReLU activation function were used (Figure 5). The input size was kept at 200*200*3. Max2D pooling layers with pool size 2*2 were stacked in between these convolutional layers. Fully connected dense layers with 64 nodes and 1 node was kept in the end. A dropout rate of 10% was applied after the first dense layer. The model was compiled with the 'binary_crossentropy' loss and 'rmsprop' optimisers.

FIGURE 5 MODEL CONFIGURATION FOR A3



3.4.3.2 IMAGE PREPROCESSING

This method requires a relatively large amount of data since all the weight layers are trained. Therefore, to increase generalisability of the model, data augmentation measures were implemented (Figure 6). We used the ImageDatagenerator real-time data augmentation method provided by Keras Neural Network API. It takes a batch of images for training and applies random transformations to each image in the batch. In contrast to traditional data augmentation measures, where original and transformed images are returned, it returns only the transformed images. The training images were rescaled to values from 0 to 1 as images with values from 0–255 entail a huge computation load due to large numbers while being processed. Width shift and height shift ranges were assigned 0.1 each. These are the fraction of width and height within which the images will be translated. The other arguments for data augmentation used were *shear_range* = *0.2*, *zoom_range* = *0.2* and *horizontal_flip* = *True*. Only rescaling was done in testing images. The number of iteration for model training was kept 20.

FIGURE6 AUGMENTED IMAGES OF WILD PIG AND BARKING DEER

SECTION IV

Results

4.1 Transfer learning with ImageNet pre-trained weights (A1)

The top five predictions from this approach were successful in identifying animal species, namely porcupine, macaque, common leopard and wild boar (Table 7). In particular, porcupine, macaque and common leopard were identified in most of the images provided, but probability was low and fluctuated according to the level of noise in the input images. Also, only ResNet50 was able to identify wild boar in images. For detecting macaque, both the models were useful, but VGG16 classified macaque images as squirrel monkey and spider monkey and ResNet50 as patas monkey. Similarly, common leopard was detected by VGG16 only. These models could not identify images of the other species.

4.2 Transfer learning as an integrated feature extractor (A2)

From our experiments, we achieved 95.71% accuracy in our 'Wild Boar and Deer' dataset, and 83.24% in 'Wild Boar and Others' dataset. Learning curves started converging around epoch 6. Confusion matrix showed less number of deer than wild boars being misclassified (Table 3). Classification report for this data set are shown in Table 4.

TABLE 3 CONFUSION MATRIX (WILD BOAR AND DEER DATASET)

		Predicted labels	
		Boar (0)	Deer (1)
Actual labels	Boar (0)	66	4
	Deer (1)	2	68

For the 'Wild Boar and Others' dataset, most of the images with wild boars were classified correctly except 4 (Table 5). In the 'Others' category, 34 out of 46 images were correctly classified. This shows that most of the images with wild boar are correctly classified. However, some of the images without them are also classified as wild boar. Classification report for this data set is shown in Table 6. Our study result with 83.24% accuracy with one species against the remaining ten still demonstrates robustness and the model offers considerable automation application potential in animal recognition and image classification. However, these measures have to be carefully checked during model deployment, especially when false positives and false negatives can be associated with serious costs.

TABLE 4 CLASSIFICATION REPORT (WILD BOAR AND DEER DATA SET)

	Precision	Recall	f1-score
Wild Boar	0.97	0.94	0.96
Deer	0.94	0.97	0.96

TABLE 5 CONFUSION MATRIX (WILD BOAR AND OTHERS DATA SET)

		Predicted labels	
		Boar (0)	Other (1)
Actual labels	Boar (0)	42	4
	Other (1)	12	34

TABLE 6 CLASSIFICATION REPORT (WILD BOAR AND OTHERS DATA SET)

	Precision	Recall	f1-score
Wild Boar	0.78	0.91	0.84
Other	0.89	0.74	0.81

TABLE 7 PREDICTIONS AND PROBABILITIES FOR THE SELECTED ANIMAL SPECIES

Species used for prediction	Top five probabilities values for species	
	VGG16 (Probability %)	**ResNet50 (Probability %)**
 Wild boar	Swing – 6 Banjo – 3 Bow – 1 Wallaby – 1 Guillotine – 1	Warthog – 45 Hog – 17 Badger – 11 Wild boar – 9 Guinea pig – 7
 Porcupine	Porcupine – 26 Echidna – 22 Badger – 16 Otter – 6 Platypus – 4	Echidna – 44 Porcupine – 33 Badger – 10 Grey fox – 6 Red fox – 3
 Macaque	Mongoose – 37 Squirrel monkey – 14 Spider monkey – 10 Baboon – 7 Grey fox – 4	Mongoose – 62, Patas monkey – 15 Baboon – 12 Wallaby – 2 Fountain – 1
 Leopard	Cheetah – 26 Bittern – 11 Red fox – 8 Meerkat – 8 Mongoose – 7	Ear – 70 Corn – 18 Fountain – 2 Peacock – 1 Mongoose – 0

Both learning curves show slight overfitting, with the second data set higher than the first (Figures 7A, 7B). The second data set with 'Others' category entailed images from all other remaining animal species except wild boar during training, unlike only two animal species in the first dataset. Additionally, the number of images of each animal species in the 'Others' category was less compared to the first dataset where only two species were considered. Due to this, the accuracy may have been lower compared to the model trained on the first dataset. Another plausible argument for this anomaly could be the presence of more than one animal species in the 'Others' category, eventually increasing 'noise' and presenting learning challenges to the model. The results and accuracy might improve if the number of images in the 'Others' category are increased as various features of animal species have to be learned by the model to provide those features as one output.

FIGURE 7A **CROSS ENTROPY LOSS AND CLASSIFICATION ACCURACY (>95.71 %)**

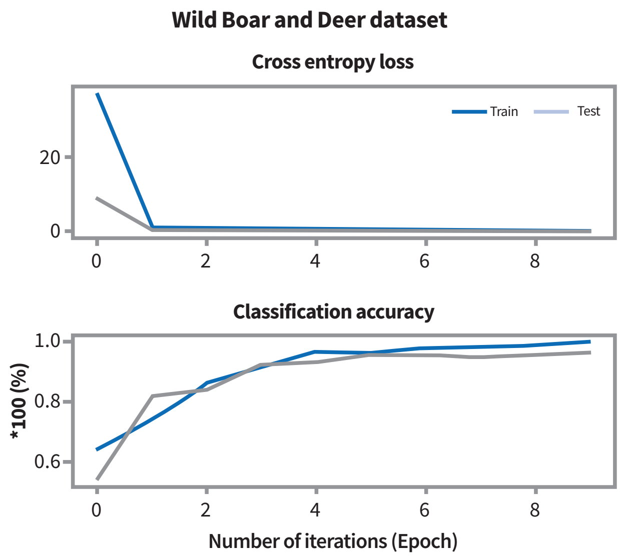
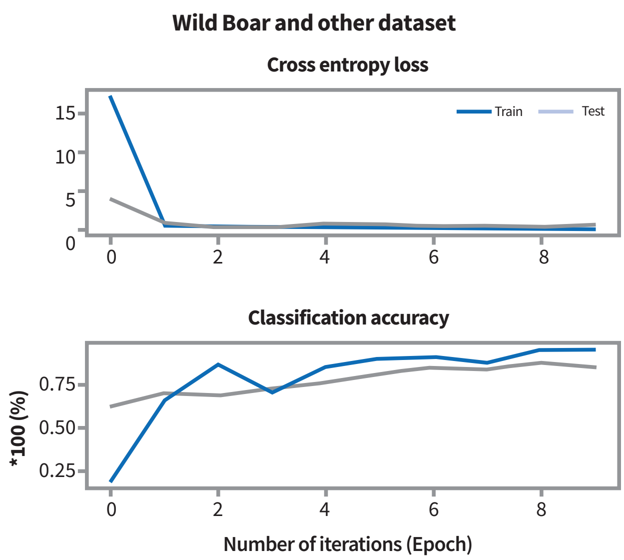


FIGURE 7B **CROSS ENTROPY LOSS AND CLASSIFICATION ACCURACY (>83%)**



4.3 Custom CNN model resembling VGG16 architecture (A3)

The accuracy in the 'Wild Boar and Deer' dataset stood at 87%, whereas only 71% accuracy was achieved in the 'Wild Boar and Others' (Figures 8A, 8B). This can be attributed to the limited amount of data used in training the model (644 and 300 in two datasets) from scratch even though data augmentation was used. The method was computationally expensive as more iterations were needed to reach a certain accuracy. Random fluctuations were seen in the model learning curves (Figures 8A, 8B). While this approach

FIGURE 8A **CROSS ENTROPY LOSS AND CLASSIFICATION ACCURACY (>87%)**

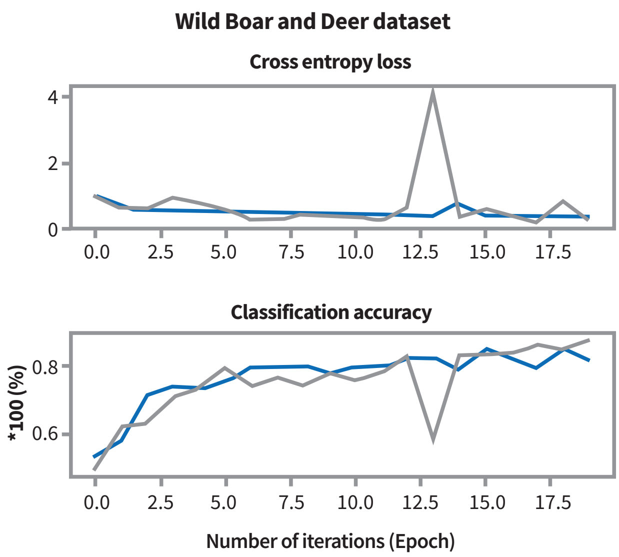
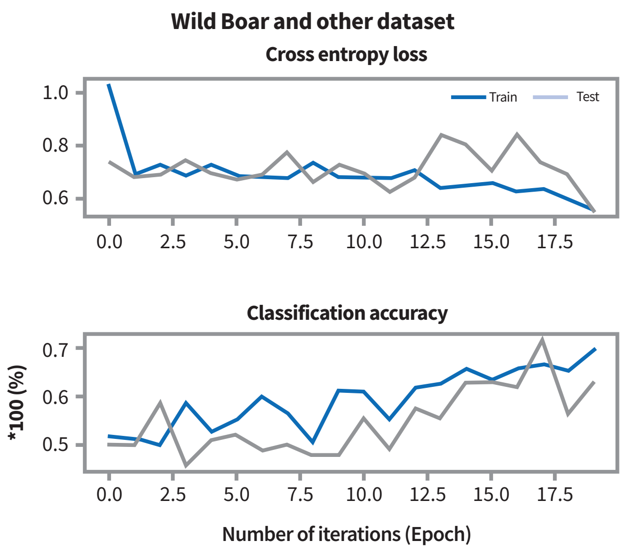


FIGURE 8B **CROSS ENTROPY LOSS AND CLASSIFICATION ACCURACY (>71%)**



seems to be a bit insufficient given the high level of accuracy achieved from the second approach for our classification tasks, it still has the potential to perform well in future with increased data size and more hyper parameter tuning.

Confusion matrix and classification report for wildboar and deer data set, and wild boar and others data set are shown in Table 8, 9, 10, and 11 respectively.

TABLE 8 CONFUSION MATRIX (WILD BOAR AND DEER DATASET)

		Predicted Labels	
		Boar (0)	**Deer (1)**
Actual labels	Boar (0)	61	9
	Deer (1)	11	59

TABLE 9 CLASSIFICATION REPORT (WILD BOAR AND DEER DATA SET)

	Precision	Recall	f1-score
Wild Boar	0.85	0.87	0.86
Deer	0.87	0.84	0.91

TABLE 10 CONFUSION MATRIX (WILD BOAR AND OTHERS DATA SET)

		Predicted labels	
		Boar (0)	**Other (1)**
Actual labels	Boar (0)	34	12
	Other (1)	14	32

TABLE 11 CLASSIFICATION REPORT (WILD BOAR AND OTHERS DATA SET)

	Precision	Recall	f1-score
Wild Boar	0.71	0.74	0.73
Other	0.73	0.71	0.72

SECTION V

Discussion

Three approaches of deep learning were explored to detect animal species using image classification models provided by Keras and Tensorflow. Our study has demonstrated results comparable to other benchmark deep learning studies carried out in animal recognition. However, direct comparison of model results and accuracy with past studies cannot be done given that the classification problems and the datasets and their size used in our study are entirely different.

5.1 Pros and cons of each approach

Many wildlife studies incorporating deep learning to classify animal species have ignored transfer learning with ImageNet pre-trained weights although the approach seems to be useful for a few selected species as has been demonstrated in our study. The reason why these three animals, namely porcupine, common leopard and monkey were classified in most images, although probability is very low, is because of their unique features. There are few species that resemble these species in ImageNet data or in general as a whole. Additionally, the method does not require any model training and testing with the lowest computational cost out of three approaches. However, the probability values for detected animal species is very low due to which in-depth analysis is required before its implementation.

While the first approach was able to detect a few animal species in images acquired from camera traps, pre-trained models used for direct classification cannot identify all animal species captured in other study locations. This is due to the fact that animal species used for training are different from place to place and there are many similar looking animals in a particular study area. And, for their classification, models need to be trained with data from that particular place. The rhesus macaque classified as 'other species of monkey' is an example in this case. Also the output of warthog and hog for input of wild boar images makes the second and third approaches ideal for long-term monitoring efforts where the occurrence of wild boar is dominant. In this regard, our second approach shows tremendous potential for scaling up in multi class classification problems, as has been demonstrated by the 'Wild Boar and Deer' dataset. The model trained with 'Wild Boar and Others' dataset also has the potential to be implemented in classifying wild boar images. In terms of computational cost, the third approach is costlier since weights have to be initialised and all layers of the model trained.

5.2 Potential for replication and scaling up in the region

Our study results show tremendous potential for implementing deep learning approaches to detect and understand the behavioural patterns of various animal species, including top conflict species in the region, in the near future. The approach has demonstrated the potential to identify occurrence of species like porcupine, wild boar, macaqueand leopard. However, more in-depth studies using different pictures of the same species should be conducted to improve robustness and accuracy. The second approach can offer us crucial first order information regarding a probable conflict zone or even alert us to the presence of these species in no time, allowing researchers and planners to carry out follow-up actions or extended measures to reduce conflict.

In the introduction section, we briefly discussed HWC conflicts in Bhutan, India and Nepal and the Kanchenjunga Landscape (KL) as a whole. The KL can particularly benefit from these approaches because of the high levels of HWC. In 2019, the impacts of crop raiding by wildlife on communities in the buffer zone of Sakteng Wildlife Sanctuary, Bhutan were studied and the results showed porcupine to be the most destructive wild animal, even more than macaque and wild pig. Another important aspect that the study discussed is the opportunity costs associated with round the clock physical crop guarding, the most widely used method to monitor and control crop raiding species. Because of this, in a season, a household is reported to experience an annual loss of about USD 268.8 calculated based on the minimum national wage rate of USD 3.6 per day. The study also showed that HWC has impacts on household food security and education. 18.2% of households were reported to have left land uncultivated due to severe crop raiding and children were reported to have dropped out of schools due to their participation in crop guarding activities. Crop guarding also entails the extra use of fuelwood, adding extra costs for conflict affected households, since porcupine and wild pig raid crops at night (Wangyel et al. 2019). Similarly, in Saephu Gewog, Wangduephodrang Dzongkhag Bhutan, wild pig is reported as a major problem animal in crop raiding followed by sambar and barking deer. Approximately, in this Gewog, potatoes worth Nu 50,000 are destroyed every year. Even with physical guarding, crop loss is about 1.5 metric tons per household. Consequently, around 1.2 ha of land per household is left fallow. It is reported that wild pigs have become uncontrollable in the area. This has been reported as a significant threat to food security. Therefore, deep learning methods combined with camera trap data along with field inputs and local knowledge can to some extent be suitable for large-scale crop monitoring and controlling crop raiding, livestock depredation, and can help reduce HWC.

5.3 Data exchange and mainstreaming of deep learning in conservation efforts

Tremendous amounts of camera trap data have been generated with various current and past wildlife monitoring programmes in the region (Wegge, Yadav, and Lamichhane 2018); Lamichhane et al. 2019; (Subba et al. 2017); (Lama et al. 2018); (Carter et al. 2015). The methods used in our study can be scaled up to automatically process these vast amounts of data so that future large-scale conservation efforts and ecological studies can be carried out in an automated manner. Given the efficiency and cost implications, and the potential applications for near real-time monitoring and conflict management decision making, it is critical that we mainstream machine learning/deep learning into conservation programmes in the region by developing a flexible data exchange mechanism between conservation projects. The more the data the better the generalisability of deep learning models in a wide array of data. A regional camera trap database and manual crowd labelling programme can be initiated with volunteers and citizen scientists to aid in labelling data for feeding into deep learning models.

Some of the key projects whose data can be used in deep learning studies in the region are:

- WWF-Nepal in collaboration with the National Trust for Nature Conservation and the Government of Nepal has deployed camera traps in various biodiversity hotspots of the country. These camera traps are particularly used for monitoring the royal Bengal tiger, one-horned rhinoceros and snow leopard. These data are also used for collecting evidence on poaching, overgrazing, HWC, unsustainable water projects and other threats causing degradation of grassland habitats. Friends of Nature, Nepal, has several conservation programmes targeted at monitoring clouded leopard, large Indian civet and other endemic mammal and bird species in Nepal.
- Similarly, a study carried out in biological corridors of protected areas and non-protected areas in Bhutan, 1858 camera traps were deployed between 2014 and 2015 to study species richness and efficacy of biological corridors in protecting mammals. As a result, 10 million photos were collected with an average of 5,368 photos per camera. Fifty six mammal species were recorded out of which 18 are threatened according to the International Union for Conservation of Nature. Data collected in a massive scale like this can be trained with machine learning to automate similar future species classification and wildlife monitoring studies in the region (Dorji, Rajaratnam, and Vernes 2019).
- The Trans-Boundary Manas Conservation Area (TraMCA) is another joint conservation initiative between India and Bhutan along with local communities and is targeted at monitoring the tiger population in the Manas National Park in Assam, India, and the Royal Manas National Park of Bhutan. The TraMCA landscape covering an area of 6500 sq. km occupies the entirety of the Manas Tiger Reserve in India, four protected areas in Bhutan and two biological corridors. The

region is considered as one of the biologically rich transboundary region and is home to more than 1500 mammal, bird and plant species. Camera trap survey was initiated in 2010-11 with the recent survey carried out in 2015 in Bansbari and Bhuyanpara Ranges of MNP – covering 240 sq. km and Umling and Manas Ranges of RMNP – covering 320 sq km (IUCN 2016).

5.4 Future work

We plan to run multi-class image classification models to identify all animal species found as we collect more data in future at the ICIMOD Knowledge Park. In addition, we want to appeal to various conservation programmes to come together for developing a database and data exchange mechanism in order to improve access to data for implementing deep learning techniques in the region. This is crucial to mainstream these approaches from a regional perspective. From there, we will pursue other possible avenues of AI and wildlife research that will take us closer to addressing more complex issues of human-wildlife conflict in the region, in addition to automating classification of animal species from camera traps, one of them being the real time monitoring of key animal species. Our future work will focus on ways to take the learning from this study and work with various national and local partners to help reduce HWC in the region.

SECTION VI

Conclusion

We demonstrated three approaches of deep learning for classifying images of wild animal species captured by trail cameras. Using different experimental settings, these methods were found to be robust and suitable for analysing images captured from trail cameras deployed in the wild.

Machine learning is sure to revolutionise the way wildlife monitoring activities are carried out in future. In the HKH region, most wildlife monitoring efforts are limited to protected areas and their periphery. There is great scope to scale up and locate more such efforts in community institutions, such as community forest user groups and local governments, to generate a more complete picture of the presence and movement of wildlife in the region. Machine learning can help us to quickly and efficiently analyse the vast amount of data that such an exercise will generate. It can also help us analyse crowd sourced images to fill data gaps in areas that are not being systematically monitored. All in all, data-driven approaches like machine learning and artificial intelligence can be a significant tool for achieving better and faster results in areas of ecological research and for managing human-wildlife interactions and conflict.

References

Acevedo, M. A., Corrada-Bravo, C. J., Corrada-Bravo, H., Villanueva-Rivera, L. J., & Aide, T. M. (2009). Automated classification of bird and amphibian calls using machine learning: A comparison of methods. *Ecological Informatics, 4*(4), 206-214. https://doi.org/10.1016/j.ecoinf.2009.06.005

Anand, S., & Radhakrishna, S. (2017). Investigating trends in human-wildlife conflict: is conflict escalation real or imagined? *Journal of Asia-Pacific Biodiversity, 10*(2), 154-161. https://doi.org/10.1016/j.japb.2017.02.003

Carter, N., Jasny, M., Gurung, B., & Liu, J. (2015). Impacts of people and tigers on leopard spatiotemporal activity patterns in a global biodiversity hotspot. *Global Ecology and Conservation, 3*, 149-162. https://doi.org/10.1016/j.gecco.2014.11.013

Chettri, N., & Gurung, J. (2017). *Human Wildlife Conflict in the Hindu Kush Himalayas: Regional Perspective*. Retrieved March 17, 2020 from https://www.researchgate.net/publication/320977346_Human_Wildlife_Conflict_in_the_Hindu_Kush_Himalayas_Regional_perspectives

Deng, J., Dong, W., Socher, R., Li, L. J., Kai, L., & Li, F.-F. (2009, 20-25 June 2009). ImageNet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition,

Dorji, S., Rajaratnam, R., & Vernes, K. (2019). Mammal richness and diversity in a Himalayan hotspot: the role of protected areas in conserving Bhutan's mammals. *Biodiversity and Conservation, 28*(12), 3277-3297. https://doi.org/10.1007/s10531-019-01821-9

Idea Connection. (2019). *Artificial Intelligence for Wildlife Conservation - Open Innovation Success Stories*. Retrieved July 28, 2021 from https://www.ideaconnection.com/open-innovation-success/Artificial-Intelligence-for-Wildlife-Conservation-00784.html

IUCN. (2016). *(Camera) Trapping Tigers across Borders*. Retrieved June 26, 2020 from https://www.iucn.org/news/species/201607/camera-trapping-tigers-across-borders

Karim, R. (2019). *Illustrated: 10 CNN Architectures*. Retrieved April 8, 2021 from https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d

Karki, S. & Udas, E. (2016). *Assessment of Forest Carbon Stock and Carbon Sequestration Rates at the ICIMOD Knowledge Park at Godavari*. ICIMOD. https://doi.org/10.53055/ICIMOD.622

KuenselOnline. (2019, Mar 8). HWC Endowment Fund Targets USD 35M. Retrieved May 22, 2020 from https://kuenselonline.com/hwc-endowment-fund-targets-usd-35m/

Lama, R. P., Ghale, T. R., Suwal, M. K., Ranabhat, R., Regmi, G. R. (2018). First photographic evidence of Snow Leopard Panthera uncia (Mammalia: Carnivora: Felidae) outside current protected areas network in Nepal Himalaya. *Journal of Threatened Taxa., 10*(8), 12086–12090. https://doi.org/10.11609/jott.3031.10.8.12086-12090

Lamichhane, B. R., Leirs, H., Persoon, G. A., Subedi, N., Dhakal, M., Oli, B. N., Reynaert, S., Sluydts, V., Pokheral, C. P., Poudyal, L. P., Malla, S., & de Iongh, H. H. (2019). Factors associated with co-occurrence of large carnivores in a human-dominated landscape. *Biodiversity and Conservation, 28*(6), 1473-1491. https://doi.org/10.1007/s10531-019-01737-4

Di Minin, E., Fink, C., Hiippala, T., & Tenkanen, H. (2019). A framework for investigating illegal wildlife trade on social media with machine learning. *Conservation Biology, 33*(1), 210-213. https://doi.org/10.1111/cobi.13104

Neurohive. (2018). VGG16 - Convolutional Network for Classification and Detection. Retrieved April 7, 2021 (https://neurohive.io/en/popular-networks/vgg16/).

Norouzzadeh, Mohammad Sadegh, Anh Nguyen, Margaret Kosmala, Alexandra Swanson, Meredith S. Palmer, Craig Packer, and Jeff Clune. 2018. "Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images with Deep Learning." *Proceedings of the National Academy of Sciences of the United States of America 115*(25):E5716–25. doi: 10.1073/pnas.1719367115.

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

NPPC, WWF. (2016). Human Wildlife Conflict: SAFE Strategy. National Plant Protection Center and World Wildlife Bhutan Program, Thimphu Bhutan. *Wildlfie Management Series*, 85.

Peterson, S., Palmer, M., Paquet, A., & Kohli, P. (2019). Using Machine Learning to Accelerate Ecological Research. Retrieved July 11, 2021 from https://deepmind.com/blog/article/using-machine-learning-to-accelerate-ecological-research

Recknagel, F. (2001). Applications of machine learning to ecological modelling. *Ecological Modelling, 146*(1), 303-310. https://doi.org/https://doi.org/10.1016/S0304-3800(01)00316-7

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision, 115*(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y

Sharma, E., Molden, D., Rahman, A., Khatiwada, Y. R., Zhang, L., Singh, S. P., Yao, T., & Wester, P. (2019). Introduction to the Hindu Kush Himalaya Assessment. In P. Wester, A. Mishra, A. Mukherji, & A. B. Shrestha (Eds.), *The Hindu Kush Himalaya Assessment: Mountains, Climate Change, Sustainability and People* (pp. 1-16). Springer International Publishing. https://doi.org/10.1007/978-3-319-92288-1_1

Sherchan, R., & Bhandari, A. (2017). Status and Trends of Human-Wildlife Conflict: A Case Study of Lelep and Yamphudin Region, Kanchenjunga Conservation Area, Taplejung, Nepal. *Conservation Science, 5*(1), 19–25. https://doi.org/10.3126/cs.v5i1.24296.

Subba, S. A., Shrestha, A. K., Thapa, K., Malla, S., Thapa, G. J., Shrestha, S., Shrestha, S., Subedi, N., Bhattarai, G. P., & Ottvall, R. (2016). Distribution of grey wolves Canis lupus lupus in the Nepalese Himalaya: implications for conservation management. *ORYX, 51*(3), 403-406. https://doi.org/10.1017/S0030605316000296

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

The Bhutanese. (2013, August 08). *Human-Wildlife Conflict*. Retrieved May 20, 2020 from https://thebhutanese.bt/human-wildlife-conflict/

Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour, 124*, 203-220. https://doi.org/10.1016/j.anbehav.2016.12.005

Wangyel, S., Tobgay, S., Dorji, K., & Wangdi, T. (2019). Impacts of crop raiding by wildlife on communities in buffer zone of Sakteng Wildlife Sanctuary, Bhutan. *International Journal of Scientific Research and Management, 7*(04), 129-135. https://doi.org/10.18535/ijsrm/v7i4.fe01

Wegge, P., Yadav, S. K., & Lamichhane, B. R. (2016). Are corridors good for tigers Panthera tigris but bad for people? An assessment of the Khata corridor in lowland Nepal. *ORYX, 52*(1), 35-45. https://doi.org/10.1017/S0030605316000661

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data, 3*(1), 9. https://doi.org/10.1186/s40537-016-0043-6

Wu, J. (2017). Introduction to Convolutional Neural Networks. *Introduction to Convolutional Neural Networks*. https://cs.nju.edu.cn/wujx/paper/CNN.pdf

About ICIMOD

The International Centre for Integrated Mountain Development (ICIMOD), is a regional knowledge development and learning centre serving the eight regional member countries of the Hindu Kush Himalaya – Afghanistan, Bangladesh, Bhutan, China, India, Myanmar, Nepal, and Pakistan – and based in Kathmandu, Nepal. Globalisation and climate change have an increasing influence on the stability of fragile mountain ecosystems and the livelihoods of mountain people. ICIMOD aims to assist mountain people to understand these changes, adapt to them, and make the most of new opportunities, while addressing upstream-downstream issues. We support regional transboundary programmes through partnership with regional partner institutions, facilitate the exchange of experience, and serve as a regional knowledge hub. We strengthen networking among regional and global centres of excellence. Overall, we are working to develop an economically and environmentally sound mountain ecosystem to improve the living standards of mountain populations and to sustain vital ecosystem services for the billions of people living downstream – now, and for the future.

REGIONAL MEMBER COUNTRIES

AFGHANISTAN BANGLADESH BHUTAN CHINA

INDIA MYANMAR NEPAL PAKISTAN

Published by
International Centre for Integrated Mountain Development (ICIMOD)
GPO Box 3226, Kathmandu, Nepal

ISBN 978-92-9115-749-5 (electronic)
https://doi.org/10.53055/ICIMOD.1016

Production team
Samuel Thomas (Senior editor)
Rachana Chettri (Editor)
Dharma R Maharjan (Graphic designer)

Citation
Shrestha, T., Martin, M.A., & Thomas, S. (2022). *Exploring the potential of deep learning for classifying camera trap data: A case study from Nepal*. Working paper. ICIMOD.

ICIMOD gratefully acknowledges the support of its core donors: the Governments of Afghanistan, Australia, Austria, Bangladesh, Bhutan, China, India, Myanmar, Nepal, Norway, Pakistan, Sweden, and Switzerland.

International Centre for Integrated Mountain Development
GPO Box 3226, Kathmandu, Nepal
T +977 1 5275222 | **E** info@icimod.org | **www.icimod.org**