Endangered Red Panda Behaviours Classification: A Comparative Evaluation of Deep Learning Models

MSc Research Project
MSc in Artificial Intelligence

Wutyi Kyi Toe
Student ID: x23194286

School of Computing
National College of Ireland

Supervisor: Prof. Rejwanul Haque

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Wutyi Kyi Toe
Student ID:	x23194286
Programme:	MSc in Artificial Intelligence
Year:	2024
Module:	MSc Research Project
Supervisor:	Prof. Rejwanul Haque
Submission Due Date:	12/12/2024
Project Title:	Endangered Red Panda Behaviours Classification: A Comparative Evaluation of Deep Learning Models
Word Count:	6770
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Wutyi Kyi Toe
Date:	29th January 2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Endangered Red Panda Behaviours Classification: A Comparative Evaluation of Deep Learning Models

Wutyi Kyi Toe
x23194286

Abstract

The significant advancements of machine learning contribute to analysing the complex patterns of data and providing insightful information to the various fields. Therefore, this research is motivated to utilise machine learning techniques for the classification of the red panda behaviours. The red panda was listed as an endangered species in 2015 due to the decline of their population up to 50% over the last 20 years. The conservation of red pandas is critical to maintaining the ecosystem balance, and understanding the behaviours of red pandas is important to engage in their conservation in wildlife and zoos. This research proposes state-of-the-art Convolutional Neural Network (CNN) models to contribute to the classification of red panda behaviours (eating, resting, and walking) using images and video footage. As the Red Panda dataset is not available publicly, the custom Red Panda dataset is created specifically for this research work. A self-trained CNN model is developed and named CNN-RedPanda; moreover, pre-trained CNN models – EfficientNetB0 and ResNet50 are implemented using the transfer learning technique. The research highlights that ResNet50 is the most accurate model with 99.27% while EfficientNetB0 is balanced in accuracy and computational efficiency. The self-trained CNN-RedPanda model achieves considerably good performance at 91.35% accuracy and can be regarded as the most lightweight model among three of them. This research work contributes to the conservation of endangered red pandas with the novel approach and optimal results of the deep learning models.

1 Introduction

Red panda (Ailurus fulgens) is the small mammal species found in Asia, especially the Himalayan area of Bhutan, China, India, Myanmar, and Nepal (Glatston and Sherpa, 2015). According to the red list of the International Union for Conservation of Nature (IUCN), the red panda species is classified as endangered due to the loss of their habitats and fragmentation because of commercial development, world climate change, human disturbance such as hunting and trapping, as well as illegal trading to border countries and poaching for medicines, pelts, and so on (Glatston and Sherpa, 2015; Ko Lin et al., 2022; Red Panda Network, 2024*b*). The population of red pandas is decreasing gradually, and only 10,000 are in the wild, as stated by the World Wildlife Fund (WWF). Therefore, monitoring their behaviours is essential for conservation efforts, not only for organisations such as the Red Panda Network but also for zoo environments.

Observation of animal behaviours is one of the most effective methods to understand the physical or non-physical processes and health of animals (Spiezio et al., 2022). According to the Red Panda Network (RPN) in Nepal, GPS collars are implemented on ten red pandas, and camera traps are employed in red panda habitats to monitor by local forest guardians (Red Panda Network, 2024*a*). Similar to RPN, Ko Lin et al. (2022) from the Fauna & Flora International organisation implements the trail cameras and conducts surveys with locals to understand the nature of red pandas in Northeast Myanmar as well as the cause of decreasing red pandas in this region. Despite actively contributing to the conservation of red pandas by these organisations, advanced technologies of artificial intelligence (AI) are still not yet implemented and widespread.

Deep learning techniques are well-known for image classification and recognition of animals such as mammals, reptiles, birds, and marine animals (Binta Islam et al., 2023; Sharma et al., 2024). Therefore, this research motivates the exploration of the power of convolutional neural networks (CNNs) to classify the behaviours (eating, resting, walking) of red pandas. The three models— two pre-trained models (EfficientNetB0 and ResNet50) with the transfer learning method and a self-trained model (named "CNN-RedPanda") from scratch—are built to train and compare in terms of accuracy and efficiency. Each model presents different approaches with their individual strengths and weaknesses by addressing the imbalanced custom Red Panda dataset due to the constraint of data availability. These models are analysed with metrics such as accuracy, precision, recall, and F1-score, as well as their training time.

1.1 Research Question

The mentioned research statement motivates the following research question:
In order to monitor the health and activity levels of the endangered species Red Panda through images and video footage, how can different CNN pre-trained models (EfficientNetB0 and ResNet50) and a self-trained CNN model (CNN-RedPanda) classify the movement patterns of red pandas in relation to eating, resting, and walking habits in terms of balancing between accuracy and efficiency?

1.2 Research Objective

To align with the research question, the following objectives are set:

1. Training the CNN models, using the transfer learning method for pre-trained models and data augmentation for both pre-trained and self-trained models.

2. Setting the same hyperparameters (such as epochs, batch size, optimiser, and so on) to all models in order to compare their accuracy and efficiency.

3. Analysing the results of the classification report and confusion matrix of each model and finding the most suitable one for future implementation with wildlife and zoos.

1.3 Contribution

The research mainly contributes to the following:

1. Custom dataset - The raw images and raw video footage are collected from Japanese zoo photographers and the Red Panda Network (Nepal) as well as recorded in

Dublin Zoo with two trail cameras to create a custom Red Panda dataset, and later this dataset will be published to be available for other students and researchers.

2. Behaviour Classification - The mentioned CNN models are trained and fine-tuned for performing the classification of red panda behaviours (eating, resting, and walking).

This paper comprises seven sections with individual subsections, respectively, to present the detailed experiment for the research. After this introduction section, related works of other research to understand more about the specific field for red panda conservation and animal classification with CNN models. Subsequently, the methodology section explains how the data is collected and transformed into the desired format, as well as how the models are selected to train and what evaluation metrics are used for this project. After that, the design specification covers the detailed architectures of EfficientNetB0, ResNet50, and CNN-RedPanda. Implementation presents the environmental setup and the implementation of the selected models. The discussion section will evaluate and analyse the results. The final section, conclusion, and future work summarise the overall findings about strengths and limitations as well as the future enhancements for this research project.

2 Related Work

The works related to this research project are categorised into two parts. First, conservation studies with traditional methods on red panda and their limitations are discussed. Secondly, classification and recognition about advancements in image classification and custom CNN architectures for wildlife monitoring are analysed.

2.1 Conservation Studies on Red Panda

For the red panda conservation in Nepal, (Bista et al., 2020) examine the economic factors and sociocultural factors that are leading to illegal poaching. This research conducts the qualitative interviews not only with local people, including traders, but also with enforcement officials alongside market analysts to investigate the reasons behind increased demand for red panda pelts. This research highlighted cultural values, economic conditions, and weak enforcement as the primary causes. Although this research relies on quantitative data analysis and statistical tests such as Pearson's chi-square, the limited availability of data makes the research difficult to make the conservation strategies scalable.

Conservation for red pandas has been a significant focus in ecological research, specifically for the vulnerability of the species due to human-induced threats and habitat degradation. In alignment with (Bista et al., 2020), (Nature Conservation Division, Department of Forests and Park Services, 2022) also performs the community interviews, geospatial mapping, and field surveys in South-Western Bhutan in order to explore habitats of red pandas and the major impacts of human activities such as deforestation and livestock holding. This research highlights the red panda conservation by highlighting the conflicts between humans and wildlife and also encouraging community-based conservation initiatives. According to the authors, the significant habitat fragmentation and low awareness among communities may result in more difficult conservation efforts. Therefore, this research makes the inclusion of local perspectives fostering the holistic understanding of threats for red pandas. However, this approach has the limitation of getting small sample data and potentially biased survey responses.

Similarly, in North-East Myanmar, (Ko Lin et al., 2022) aim to utilise the historical records and current knowledge on the conservation status of red panda. This research implements the camera trapping and interviews in villages around Mt. Imawbum from 2010 to 2018 as well as integrates GIS mapping, population sampling, and habitat analysis to examine critical habitats and threats such as agricultural expansion and logging. This study emphasises how protected areas are important and also should be established for the conservation of red pandas. Although this research provides significant contributions to mitigate the habitat loss, similar to (Bista et al., 2020) and (Nature Conservation Division, Department of Forests and Park Services, 2022), it also has certain limitations of its geographic scope and irregular protecting patterns for habitats of red pandas.

Although the ecologists, biologists, and researchers have actively involved in the conservation of red pandas with various methods, such as surveying and installing camera traps in different regions, these methods do not effectively work for this endangered species and require significant manual work for them. With the usefulness of artificial intelligence, especially deep learning techniques, the animal classification can be performed easier and the manual work can be lessened, increasing accuracy for the classification and identification of animal species and their behavioural activities. By identifying their behaviours, the illness, stress, and potential dangers for the animals can also be detected using noninvasive methods with the advancement of AI technology (Spiezio et al., 2022).

2.2 Advances in Image Classification and Custom CNNs Architectures for Wildlife Monitoring

Classification and recognition for animals and behavioural patterns are essential for monitoring biodiversity and contributing to wildlife conservation. Many research experiments have been accomplished, and their papers have been published on animal species and behavioural classification with deep learning algorithms, particularly through convolutional neural networks (CNNs). This subsection discusses the distinct advancements and usefulness of CNNs in image classification tasks using specialised techniques and innovations in custom architectures for wildlife conservation.

2.2.1 Advancements in Pre-trained CNNs Architecture with Transfer Learning for Image Recognition and Classification

Firstly, the usefulness of deep learning with transfer learning technique is discussed by (Zhang et al., 2020) for the classification of remote sensing images. The authors utilise segmentation techniques to prevent overfitting of the model. Later, this research applies transfer learning to the pre-trained EfficientNet models. According to the authors, the proposed method is effective and achieves very optimistic results on different datasets, as well as being lightweight and less consuming for computational resources, although their method does not always achieve the highest accuracy results.

After (Zhang et al., 2020), (Mekruksavanich et al., 2022) present animal activity recognition (AAR) using a deep learning pre-trained model, deep residual networks (ResNet) to enhance the accuracy and effectiveness of observing animal behaviours with wearable sensors. The authors share similarities with (Zhang et al., 2020) for the data preprocessing steps such as segmentation and normalisation, as well as utilising the transfer learning method to the pre-trained ResNet model. With the highest accuracy of 96.46% and 98.40% for first and second sensors, (Mekruksavanich et al., 2022)'s proposed model demonstrates that it can work with two different sensors effectively while achieving high accuracy. However, the authors highlight that they will examine different target data to achieve the robust model.

Subsequently, the authors (Manivannan and Venkateswaran, 2023) apply the transfer learning with a pre-trained deep CNN architecture, Inception-ResNet-V2, to identify 120 dog breeds. The authors aim to achieve the advantages of Inception and ResNet architectures with ensemble learning. After the preprocessing and segmentation of data, the hyperparameters are fine-tuned for the proposed model. Using this approach, it benefits on classification tasks with high accuracy and speed. Moreover, the result of the proposed model is 95.03% on training data and 92.92% on validation. According to the authors, their fine-tuned model aids for dog owners to identify when their lost dogs, preventing fraud for dog sales as well as assisting veterinarians. Although the model achieves high accuracy over 90%, the testing model drops up to 88.92% of accuracy; therefore, the model still indicates rooms for improvement with more dog breeds.

According to (Binta Islam et al., 2023), their research demonstrates the recognition of animal species with the use of deep convolutional neural networks, which enhance classification accuracy from ecological camera trap images. Hence, pre-trained ResNet models prove high accuracy in (Mekruksavanich et al., 2022) and (Manivannan and Venkateswaran, 2023) for classification and recognition of sheep and dogs; (Binta Islam et al., 2023) experiment VGG16 and ResNet50 as their selected pre-trained models, as well as the authors construct CNN-1 as their self-trained model. Their research is particularly for the reptile animals, snakes, lizards, and toads; these data are obtained from five regions of Texas. After balancing, preprocessing, and augmenting the imbalanced dataset, the multi-classification results are comparatively good. Therefore, using transfer learning for all models, VGG16 and ResNet50 achieve 87% and 86% respectively, while the self-trained CNN-1 attains 72%. This research exhibits high achievement in both pre-trained and self-trained models, and VGG16 outperforms among three models. Despite performing high accuracy for recognition of specific reptile animals, no experiment is indicated for mammals, especially for endangered species.

While (Manivannan and Venkateswaran, 2023)'s research emphasises the classification of dog breeds, (Cahyo et al., 2023) also present cat breed classification using fine-tuned CNNs, particularly Xception with transfer learning. The Oxford-IIIT Pet Dataset contains 12 cat breeds and is utilised for this approach. The authors demonstrate two scenarios: CNNs with transfer learning and CNNs with transfer learning and fine tuning. Similar to prior research papers, (Cahyo et al., 2023) apply the same preprocessing and augmentation for images. With the high accuracy of 92.5% for transfer learning with fine-tuned CNNs, their approach is faster and more cost-effective than traditional DNA tests to help cat owners identify cat breeds. On the other hand, their model should be trained with all possible breeds to prevent overfitting.

Similar to Manivannan and Venkateswaran (2023),Mekruksavanich et al. (2022) and Cahyo et al. (2023), the authors (Valarmathi et al., 2023) propose a comparative analysis of hybrid deep learning algorithms utilising dog breed images. The authors employ the deep learning algorithms including Xception, VGG19, NASNetMobile, EfficientNetV2M, ResNet152V2, and hybrid models combining Inception-v3 & Xception and EfficientNetV2M, NASNetMobile, Inception & Xception12. Moreover, transfer learning and data augmentation are also applied in their pre-trained models. During these experiments, hybrid models significantly outperform a single model, specifically the validation accuracy of 92.4% for the hybrid of Inception-v3 and Xception, compared to 71.63% of accuracy of ResNet101. Although their models achieve high accuracy except VGG19 which accuracy is 55%, the computational resources are constraint for training these models effectively.

The motivation for employing the ensemble learning approach from (Manivannan and Venkateswaran, 2023), (Oion et al., 2023) propose the automated system for classifying marine animals using CNNs, especially the ensemble technique with EfficientNet B0, EfficientNet B3, and EfficientNet B5. Their approach aims to identify and categorise different marine species using visual characteristics. The authors obtain a large dataset with many marine species that have different shapes, colours, and sizes. They also utilise the data augmentation methods to achieve a robust model. Their proposed model accomplishes up to 87.99% for the classification task. However, the authors do not mention the computational power for their proposed system, which can make other researchers difficult to contribute for utilising their system in marine environments.

Later, Neeli et al. (2023) demonstrate the innovative bird species detecting model using CNN and the EfficientNet-B0 architecture. The purpose of this research is to identify the bird species with a precise and efficient model to conserve wildlife and monitor biodiversity. The data with 84,635 training images of 525 categories are obtained from Kaggle, and these images are standardised and augmented for better results. Moreover, Adam Optimiser and early stopping are also applied during training. The authors state that their model achieves state-of-the-art performance with the test accuracy from 89% to 92.0% and exhibits the balance between computational efficiency and high accuracy despite having some species with similar visual appearances or fewer occurrences in the dataset.

On the other hand, the research (Sharma et al., 2024) proposes utilising deep learning algorithms, particularly LeNet-5, AlexNet, VGG16, ResNet50, Inception V3, Inception, and ResnetV2, with the data augmentation and dropout to enhance the performance of the model and prevent overfitting. The authors aim this research to identify the cause of wild animal attacks and disappearances in villages or farms that are close to forests and mountains. They obtain the dataset from Kaggle, including the images of 90 different animals. With their proposed model, not only farmers and other people who live near the area of forests and mountains by recognising animals easily, but also rare or endangered animals can be classified potentially in the creation of an alert system. According to the authors, InceptionV3 performs well with the validation accuracy of 88.26% and with its F1 score of 1.31, but some models show poor performances in validation accuracy due to the small dataset.

2.2.2 Self-trained CNNs Architectures for Animal Classification Tasks

According to Song et al. (2022), the authors apply novel CNN architecture for the purpose of classification of 10 different animals. This research focuses on analysing the parameters of the algorithm to achieve a better model and application for the classification task. 28,000 high-quality images are preprocessed (such as resizing to 100 x 100 pixels), augmented, and organised into 10 categories. Their CNN model consists of multiple convolutional layers with various filter sizes, pooling, and dense layers. After the model is well constructed by fine-tuning the hypermeters, the final model achieves over 85% and 75% accuracy on training and validation. According to the authors, their final model has the drawback of being significantly time-consuming, as the model takes 20 minutes for each epoch, for a total of 60 epochs.

Priya et al. (2023) also propose Deep Convolutional Neural Networks (DCNN) to classify animal species through input images to aid in wildlife conservation and behaviour studies. The authors employ preprocessing techniques such as normalisation and augmentation, feature extraction, and transfer learning. The proposed system benefits biodiversity monitoring for a robust model with a high accuracy rate of 98% in classifying the animal species. Despite being more effective compared to traditional analysis methods, their model can be potentially overfitting and will not be utilised for endangered species such as red panda or sea otter due to the limitation of a small dataset with specific animals.

The classification of four African wildlife species, including rhino, zebra, elephant, and buffalo, is experimented by (Li, 2024). This research focuses on examining how different kernel sizes of CNN architecture affect the performance of the proposed model. The dataset of African wildlife from Kaggle consists of 1508 images divided into four stated animal categories. According to the author, the kernel sizes of the proposed model are modified to (2,2), (3,3), and (4,4) to examine the impact on accuracy. Although this experiment obtains 57.57% for (3,3) kernel size, it still has room to improve the accuracy of the proposed model to identify efficiently.

As the prior research works proved, the pre-trained CNN models such as ResNet and EfficientNet with the transfer learning method outperformed to classify and analyse the different animal species and also for their behaviours in the purpose of animals and their habitat conservation. This research also motivates and aims to employ state-of-the-art CNN models such as ResNet50, EfficientNetB0, and a self-trained CNN model to provide a long-term solution about the automated classification and monitoring of red panda behaviours for zoos and wildlife conservation efforts. Using CNN pre-trained and self-trained models, observing endangered species and their specific behaviours are crucial actions for researchers for their conservation purposes and to tackle the existing and potential threats for red pandas and other endangered species in the future.

3 Methodology

This section presents the collected data, how this data is preprocessed, and how the transformation is performed. After that, the modelling part is explained.

3.1 Custom Data Collection

One of the significant aspects of this project is the creation of a custom dataset of red panda images. The various images and video footage of red pandas are collected, and the custom dataset is created, as there is no public dataset available for red pandas. Unlike other public datasets, the video footage is recorded at the Dublin Zoo with two camera traps. Secondly, the recorded videos and images of red pandas are provided by Japanese zoo photographers. Moreover, Red Panda Network Nepal also shares some valuable images of red pandas and insightful information about red pandas. The total of collected images and video footage is about 1005, and each video footage is an average of 30 seconds in length. However, red pandas are occasionally moving out of the views from cameras, as well as the environmental factors such as weather, lighting, and obstructions (e.g., branches and walls), which are the reasons to reduce the quality and numbers of video footage.

Figure 1: Camera traps setup in Dublin Zoo used for collecting data

3.2 Data Preprocessing

Firstly, the collected images and video footage of red pandas are carefully analysed to remove the poor-quality images and videos as well as the video contents that mainly involved human beings rather than red pandas or no red panda footage. After cleaning these data, the Python script is utilised for extracting the frames of each video footage. As red pandas are mostly sleeping, slow-moving, and hibernating most of their time, the frames are extracted every two seconds per frame. Subsequently, the structural similarity (SSIM) index of 0.97 is applied for checking the similarity percentage of the extracted frames from video footage (Syed Zaini et al., 2019). Later, the basic preprocessing step, cropping unnecessary parts from the images, is proceeded (Cahyo et al., 2023) before resizing the images. Finally, the images are manually labelled to be set into three categories: eating, resting, and walking, and the total images for the custom Red Panda dataset are 4,801, but there is a slight imbalance, as the "eating" class contains more images than "walking" and "resting".

3.3 Data Transformation

During this step, the preprocessed Red Panda dataset is transformed into a well-formatted dataset that is utilised for deep learning models. Firstly, the images are resized to 256 x 256 to standardise the input size. After resizing, data augmentation methods such as rescaling pixel values, shear transformations, zooming randomly, flipping images horizontally, rotating images, and shifting the width and height of the images, adjusting brightness, shifting RGB channels randomly, and filling pixel outsides boundaries with nearest values so that the dataset has more variation and is robust as well as prevents

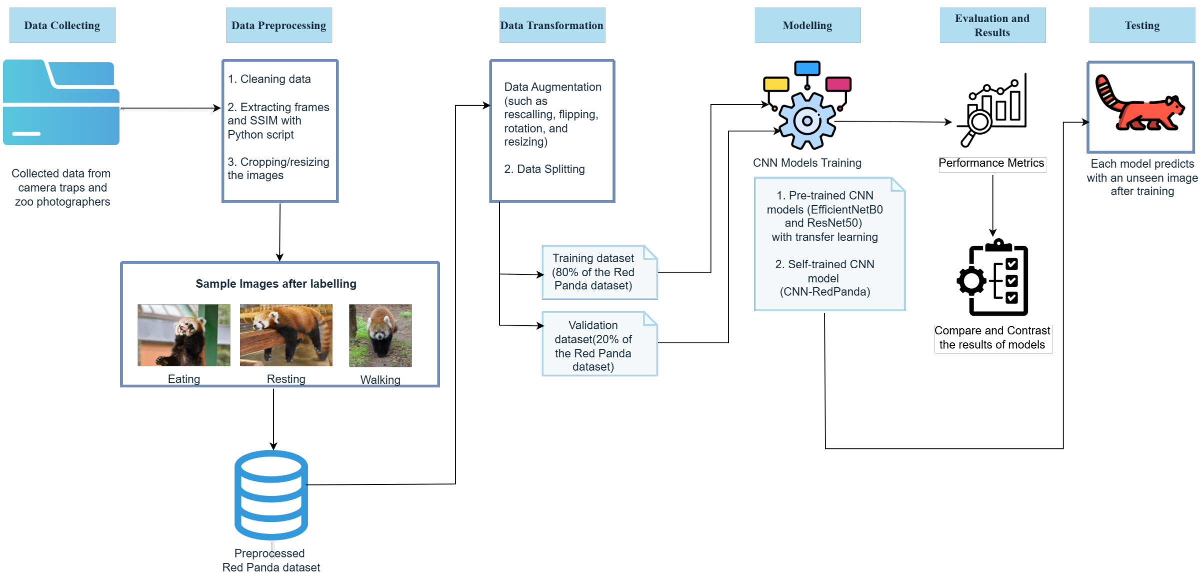

Figure 2: Research Methodology for classification of the behaviours of Red Pandas

overfitting (Cahyo et al., 2023; Neeli et al., 2023; Sharma et al., 2024). The following Figure 3 is a set of sample images extracted from the Red Panda dataset.

3.4 Data Modelling

This process involves how the data are divided for training and validation, which CNN models are selected, as well as how the trained models are compared to one another in terms of accuracy and efficiency.

3.4.1 Data Splitting

In this research, the data is split into training and validation as 80% and 20% from the total number of 4,801 images. By doing this, overfitting can be mitigated for the small dataset, such as the Red Panda custom dataset.

3.4.2 Model Selection

EfficientNetB0 and ResNet50 are chosen for their proven accuracy in image classification tasks (Binta Islam et al., 2023; Neeli et al., 2023; Oion et al., 2023; Sharma et al., 2024); additionally, they are able to perform well with small datasets when transfer learning is employed to adapt them to the red panda behaviour classification task. Having an intuitiveness of building a CNN model from scratch (Song et al., 2022; Binta Islam et al., 2023; Li, 2024), a self-trained CNN model named CNN-RedPanda is also developed.

3.4.3 Model Comparison

To evaluate the performances of the models, EfficientNetB0, ResNet50, and the CNN-RedPanda will be compared for the accuracy of the classification behaviours of red pandas as well as their efficiency for the custom dataset during training. In order to be consistent, the same hyperparameters, which are the optimiser, learning rate, batch size, epochs, applying early stopping, and loss function, are applied to all chosen models.

Figure 3: Three different behaviours of red pandas

3.5 Evaluation Metrics

Not only pre-trained (EfficientNetB0 and ResNet50) models but also the self-trained model (CNN-RedPanda) is compared using a combination of performance metrics to evaluate the classification ability of each model. The accuracy along with the efficiency of trained models is the main measurement for this research. In addition, the classification report with precision, recall, and F1-score, as well as the confusion matrix, are presented. Accuracy and loss curves of training and validation are also plotted.

4 Design Specification

This section covers the architecture designs of EfficientNetB0 and ResNet50 and CNN-RedPanda models. EfficientNetB0 and ResNet50 are built with the same architecture design using the transfer learning (Neeli et al., 2023; Sharma et al., 2024). All three models are aimed at comparing the accuracy of the classification of the three behaviours (eating, resting, and walking) of red pandas. Therefore, these models will be applied with the same algorithm-based hyperparameters, such as learning rate and epochs. There are three subsections: the first is the design architecture of EfficientNetB0 and ResNet50, followed by the design architecture of CNN-RedPanda, and finally, defining the same hyperparameters for all models.

4.1 Design architectures of EfficientNetB0 and ResNet50

The EfficientNetB0 is chosen for its performance and efficiency on image classification tasks while leveraging pretrained ImageNet weights (Neeli et al., 2023), it is greatly versatile and more adaptive than other pre-trained models. With transfer learning, the model can be fine-tuned for the Red Panda dataset by retrieving its general and specific features. As a result, the model can perform to classify the behaviours of red pandas. According to Binta Islam et al. (2023) and Sharma et al. (2024), ResNet50 is also well-known for image classification tasks, as its 50-layered architecture allows extracting high features and training with a deeper network. ResNet50 is useful for being versatile and providing pre-trained weights on ImageNet; therefore, it can perform model training with high capability for red panda behaviour classification while utilising transfer learning.

The input images are resized as 256 x 256 x 3 to be compatible with both model architectures. The initial layers of both pre-trained models are frozen except for the last 10 layers so that the models allow fine-tuning on the custom Red Panda dataset. According to Binta Islam et al. (2023), both models are enabled to be trained for the specific behaviour classification for red pandas with the transfer learning approach. For both models, the classification heads are added to customise them for the classification of the behaviours of red pandas. Therefore, the details of the additional layers are a GlobalAveragePooling2D layer to reduce the spatial dimensions of the feature maps, followed by a Dense layer that has 256 units with ReLU activation and L2 regularisation, as L2 = 0.01 is utilised to prevent overfitting (Binta Islam et al., 2023). Batch normalisation is also applied in order to accelerate and stabilise learning during the training (Binta Islam et al., 2023). Afterwards, a dropout layer with a rate of 0.5 is added to the architecture to tackle overfitting. The output layer consists of a dense layer that has three units, representing the three behaviour classes of red pandas, along with a Softmax activation function.

4.2 Design architecture of CNN-RedPanda

The self-trained CNN model, known as CNN-RedPanda, is also developed to classify the red panda behaviours. The architecture of CNN-RedPanda includes the convolutional and pooling layers to extract hierarchical features of the input images. The model initialises with a Conv2D layer, which has 32 filters of size 3 x 3, followed by a MaxPooling2D layer to reduce spatial dimensions with the purpose of managing computational complexity (Sharma et al., 2024). Thereafter, the second and third layers are also applied in the same pattern with the filters of 64 and 128. Each stated convolutional layer uses the ReLU activation function to introduce non-linearity (Binta Islam et al., 2023).

Subsequently, the Flatten layer performs converting the feature maps into a 1D vector for the input to the fully connected layers, followed by the Dense layer, which contains 256 units with ReLU activation (Song et al., 2022). Later, the Dropout layer is added into this architecture with a 0.5 rate to prevent overfitting by deactivating random neurones during the training of the model (Binta Islam et al., 2023). Afterwards, the final (output) layer, a Dense layer with 3 units and Softmax activation function for the classification of red panda behaviours.

4.3 Hyperparameters for EfficientNetB0, ResNet50, and CNN-RedPanda models

Same as Binta Islam et al. (2023), the Adam optimiser is applied with the learning rate of 1×10^{-4} using SparseCategorical Crossentropy as the loss function after the architecture of each model is customised and set up. As the main purpose of the research project is to classify correctly the behaviours of red pandas, "accuracy" is chosen as a main evaluation metric. Due to the imbalance of classes in the dataset, class weights are dynamically computed with the training dataset (Bakırarar and ELHAN, 2023). Moreover, early stopping is also applied to monitor the validation loss and stop the training when no improvement is observed over three consecutive epochs so that the best prior model weights can be restored (Binta Islam et al., 2023). Therefore, the combination of a robust architecture of each model with well-defined hyperparameters ensures the generalisation of unseen data and the classification of the behaviours of red pandas with high accuracy.

5 Implementation

This section presents the environmental setup and end-to-end implementation of pre-trained CNN models (EfficientNetB0, ResNet50) and self-trained CNN model.

5.1 Configurations and Setup

This project mainly utilises Python version 3.10.12 to build the desired classification models, along with TensorFlow version 2.17.1 and Keras version 3.5.0, ensuring compatibility with the chosen architectures and necessary libraries such as sklearn, Numpy, Matplotlib, and so on. Microsoft Visual Code IDE is utilised to extract images from video footage. All models are trained and tested in the Google Colab cloud environment using Google Drive as the main data storage for raw data and the preprocessed dataset. During model training, the T4 GPU has 15 GB of RAM, which enables efficient computational power as well as saves the training time.

5.2 Dataset Preparation

Both pre-trained and self-trained models use the same dataset; the dataset is prepared with the following steps before models are structured and trained.

1. A custom Python script using OpenCV is developed to extract unique frames from uploaded video footage by calculating the Structural Similarity Index (SSIM) to ensure the unique frames are extracted for custom dataset creation.

2. Using a Python script that utilises the Regions of Interest (ROI) selector of OpenCV, the extracted images are cropped to focus on the behaviours of red pandas and to remove unnecessary parts in each frame.

3. The images are labelled into three different classes (eating, resting, and walking) and prepared a dataset.

5.3 Implementation Steps of EfficientNetB0 and ResNet50

Since both models are built with the same architectural designs, the following implementation is applied while training both models individually.

1. The preprocessed dataset is split into 3841 and 960 to train and validate, respectively.

2. The data augmentation is applied to the training dataset so that the data will be in various transformations as well as overfitting can be overcome.

3. During the training process with custom layers for each pre-trained model, the Adam optimiser with 1×10^{-4} learning rate, batch size 32, and Sparse Categorical Crossentropy loss are applied.

4. The class weights are also computed to address the imbalance of classes in the dataset by utilising compute_class_weight from sklearn.

5. ResNet50 finds the best validation accuracy at 99.27% at 20 epochs, while EfficientNetB0 achieves 98.85% for validation accuracy at 20 epochs.

6. Along with accuracy, F1-score, precision, and recall on the classification report and visualising the confusion matrix as well as the accuracy and loss of both models are also plotted.

7. Finally, unseen images are used for testing the EfficientNetB0 and ResNet50 models.

5.4 Implementation Steps of CNN-RedPanda

1. Same as EfficientNetB0 and ResNet50, the preprocessed dataset is split into 3841 and 960 to train and validate, respectively.

2. The data augmentation techniques are utilised on the training dataset to be increased and applied to various transformations so that the overfitting of the model can be mitigated.

3. The Adam optimiser with 1×10^{-4} learning rate with batch size 32 is set, and the loss function as Sparse Categorical Crossentropy is applied.

4. The class weights are also computed to address the imbalance of classes in the dataset by utilising the compute_class_weight function from the sklearn library.

5. Although the model is trained with 20 epochs, the early stopping is triggered at 10 epochs while monitoring the validation loss; the best weights of the models are restored at epochs 9, with a validation accuracy of 91.35%.

6. The classification report, confusion matrix, as well as model accuracy and loss are also plotted.

7. Finally, an unseen image is used for testing the CNN-RedPanda model.

6 Evaluation

The performance of the EfficientNetB0, ResNet50, and CNN-RedPanda models is classified based on accuracy with precision, recall, and F1-score. The following subsections present the evaluation of each model according to its training and validation accuracy and loss curves, as well as the confusion matrix and classification report.

6.1 EfficientNetB0

The EfficientNetB0 model demonstrates exceptional performance in the classification of red panda behaviours, with a validation accuracy of 98.85%. This model has a stable improvement at 20 epochs by exhibiting strong convergence with reducing loss from 4.4908 to 0.6984 while validating with unseen 20% data. The model proves that it is highly effective and good at generalisation while presenting the high precision, recall, and F1-score for all three behaviours, especially F1-scores achieve 0.99 for all classes (eating, resting, and walking). Moreover, the training and validation accuracy curves indicate the consistency of the model, learning with no overfitting while the loss curves gradually decline across epochs. Despite being relatively imbalanced in the dataset, customising the EfficientNetB0 model with transfer learning achieves high accuracy for classification with minimal misclassifications. Therefore, EfficientNetB0 is robust and reliable, along with data augmentation and fine-tuning the hyperparameter. Figure 4 represents the accuracy and loss curves of the model.

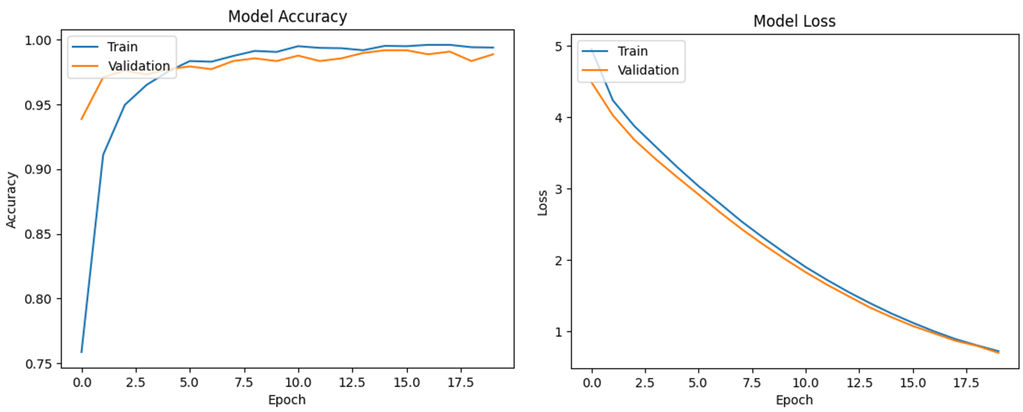

Figure 4: Accuracy and Loss Curves of EfficientNetB0

6.2 ResNet50

The ResNet50 model also demonstrates outstanding results in the classification of red panda behaviours while achieving an overall accuracy of 99.27% with 20 epochs. Same as EfficientNetB0, this model also proves for smooth convergence as the validation loss is decreasing up to 4.3259 from 0.2803. Its performance metric along with precision, recall, and F1-score, while F1-scores of eating, resting, and walking classes achieve 0.99 accordingly. The scores reflect that customising ResNet50 with transfer learning exhibits almost perfect classification across all behaviour classes (eating, resting, and walking) while encountering minor misclassification. This significant classification shows that the deep residual networks of ResNet50 can effectively handle the complex features with the usefulness of data augmentation and L2 regularisation.Figure 5 represents the accuracy and loss curves of the model during training and validation.

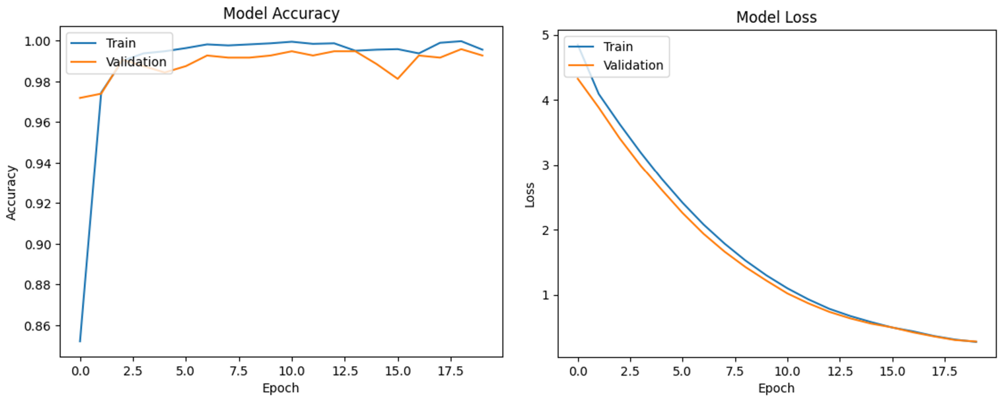

Figure 5: Accuracy and Loss Curves of ResNet50

6.3 CNN-RedPanda

The last model is the self-trained CNN model, which achieves the validation accuracy of up to 91.35%, proving that the model is capable of classifying the different behaviours of red pandas despite being built from scratch with a simple architecture compared to the other pre-trained models (EfficientNetB0 and ResNet50). It performs reasonably well while presenting over 0.90 for all precision, recall, and F1-score except recall for the eating and walking classes, which are 0.87 and 0.86, while F1-scores for eating, resting, and walking achieve 0.91, 0.93, and 0.90, respectively. The confusion matrix reveals the misclassified numbers of each class are slightly higher than pre-trained models. The process of model training is 0.3226 validation loss with the best epoch 9. Although its architecture is lightweight and efficient, its loss graph shows that CNN-RedPanda still has more room to improve its performance. Figure 6 represents the accuracy and loss curves of the CNN-RedPanda model.

6.4 Discussion

The evaluation of three models—EfficientNetB0, ResNet50, and CNN-RedPanda—presents their strengths, weaknesses, and suitability for red panda behaviour classification (eating, resting, and walking). The performance metrics of each model measure accuracy, precision, recall, and F1-score, which are analysed to assess how the models can handle the imbalanced dataset.

According to the classification reports of the models, the EfficientNetB0 achieves 98.85% validation accuracy with balanced performance in all classes. It exhibits the efficiency and the capability of strong generalisation while utilising transfer learning and regularisation, despite having the imbalanced walking class. Secondly, the ResNet50 also achieves

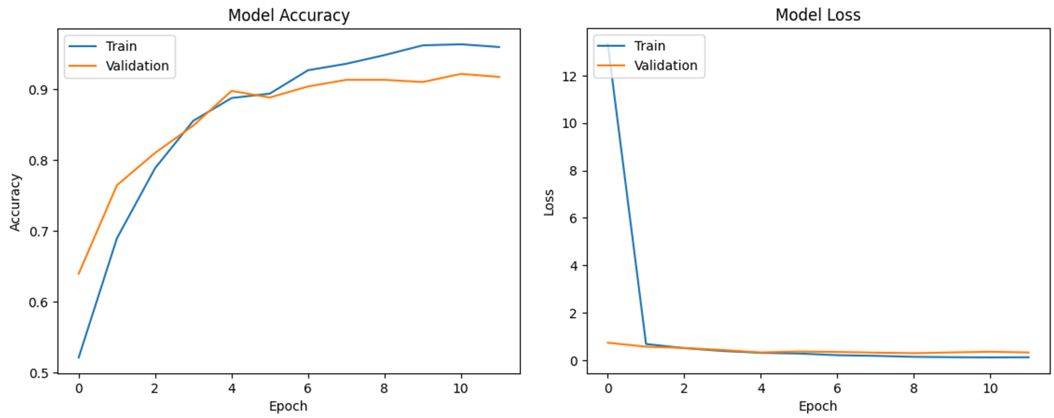

Figure 6: Accuracy and Loss Curves of CNN-RedPanda

the accuracy of 99.27% with near-perfect classification. On the other hand, it can consume more computational power than the other two models, but this model shows its consistent accuracy for generalisation and robust feature extraction on complex images. Finally, the CNN-RedPanda model demonstrates that it can perform moderate classification with environments for constrained computational resources, although its accuracy is not as high as the two pre-trained models for the custom Red Panda dataset. According to the confusion matrix, these models can be improved to reduce the misclassification by managing the imbalanced data with more augmentation, as well as oversampling with synthetic data generation (Bao, 2023) or collecting more data from zoos and wildlife organisations. The overall comparison of three models according to their classification reports and training times is displayed in Table 1.

Metric	**EfficientNetB0**	**ResNet50**	**CNN-RedPanda**
Accuracy(Validation)	98.85%	99.27%	91.35%
Precision(Average)	0.99	0.99	0.93
Recall(Average)	0.99	0.99	0.92
F1-Score(Average)	0.99	0.99	0.93
Training Time per Epoch(Average)	~91.8s	~111s	~85.8s
Strengths	Efficient for generalisation and scaling	Best accuracy with robust feature extraction	Lightweight, efficient for limited resources
Weaknesses	Slightly slower in training	Computationally intensive	Lower precision and recall for eating and walking classes

Table 1: Comparison of EfficientNetB0, ResNet50, and CNN-RedPanda Models for Red Panda Behaviour Classification

7 Conclusion and Future Work

This research compares the EfficientNetB0, ResNet50, and CNN-RedPanda for the classification of the behaviours of red pandas by evaluating their accuracy, efficiency, and overall performance along with their advantages and disadvantages. Among the three models, ResNet50 outperforms with the highest accuracy as 99.27% and is most reliable for classification, but its drawback is consuming more computational power and taking more time for training. So, the model is appropriate for environments that focus only on accuracy. Secondly, the EfficientNetB0 achieves 98.85% as the significant accuracy with balance in computational efficiency, as it takes ~91.8s per epoch during training, while CNN-RedPanda also obtains 91.35% as a reasonable accuracy with a simple architecture and less training time, so CNN-RedPanda is the most lightweighted model among them and suitable for the environments that do not require accuracy as a top priority. Overall, EfficientNetB0 is chosen to perform the red panda behaviour classification due to its balancing ability for accuracy and efficiency despite having an imbalanced class in the dataset. This research will enhance the EfficientNetB0 model with oversampling or ensemble learning methods, and more Red Panda data with other behaviours will be collected to train with EfficientNetB0. Later, a web application will be developed to allow users to upload images and classify the behaviours so that researchers can utilise this application either in wildlife or in zoo environments for red panda conservation.

8 Acknowledgement

I would like to express my wholehearted gratitude to Professor Rejwanul Haque for his guidance and unwavering support during my practicum part 2. I would like to extend my heartfelt thanks to Dublin Zoo, especially Andrew, Eddie, and Louise for their kind help and allowing me to implement trail cameras. I am sincerely thankful to Kiarino-san, Kaki-san, Darianui-san, Rostem-san, and Tao-san who shared their red panda video footage from Japanese zoos for my research. Additionally, I am grateful to explore insightful articles and images about red pandas from the Red Panda Network (Nepal), specifically from Mr. Sonam Tashi Lama as well as Ko Aung Ko Lin (Fauna & Flora) and Ye Htet Lwin for sharing with me about red panda conservation in Nepal and Myanmar. Lastly, I am deeply appreciative to Su Yi, Min Ko, and the international red panda community, whose encouragements are truly supportive to continue my current and further research for red pandas and wildlife conservation with AI technology.

References

Bakırarar, B. and ELHAN, A. (2023), 'Class weighting technique to deal with imbalanced class problem in machine learning: Methodological research', *Turkiye Klinikleri Journal of Biostatistics* **15**, 19–29. `https://doi.org/10.5336/biostatic.2022-93961`.

Bao, C. (2023), 'Image classification under class-imbalanced situation', *Highlights in Science, Engineering and Technology* **39**, 465–474. `https://doi.org/10.54097/hset.v39i.6570`.

Binta Islam, S., Valles, D., Hibbitts, T. J., Ryberg, W. A., Walkup, D. K. and Forstner, M. R. J. (2023), 'Animal species recognition with deep convolutional neural networks from ecological camera trap images', *Animals* **13**(9). `https://doi.org/10.3390/ani13091526`.

Bista, D., Baxter, G. and Murray, P. (2020), 'What is driving the increased demand for red panda pelts?', *Human Dimensions of Wildlife* **25**, 1–15. `https://doi.org/10.1080/10871209.2020.1728788`.

Cahyo, D. D. N., Sunyoto, A. and Ariatmanto, D. (2023), Transfer learning and fine-tuning effect analysis on classification of cat breeds using a convolutional neural network, *in* '2023 6th International Conference on Information and Communications Technology (ICOIACT)'. Yogyakarta, Indonesia, 10-10 November 2023, pp. 488-493. `https://doi.org/10.1109/ICOIACT59844.2023.10455771`.

Glatston, A., W. F. Z. T. and Sherpa, A. (2015), 'Ailurus fulgens, the iucn red list of threatened species'. Available at: https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T714A45195924.en [Accessed 10 September 2024].

Ko Lin, A., Lwin, N., Aung, S., Oo, W., Lum, L. and Grindley, M. (2022), *The conservation status of red panda in north-east Myanmar*, pp. 475–488. Available at: https://doi.org/10.1016/B978-0-12-823753-3.00023-5 [Accessed 12 September 2024].

Li, K. (2024), 'Africa wildlife prediction based on custom convolutional neural network', *Applied and Computational Engineering* **39**, 166–171. https://doi.org/10.54254/2755-2721/39/20230595.

Manivannan, S. and Venkateswaran, N. (2023), Dog breed classification using inception-resnet-v2, *in* '2023 International Conference for Advancement in Technology (ICONAT)'. Goa, India, 24-26 January 2023, pp. 1-5. https://doi.org/10.1109/ICONAT57137.2023.10080065.

Mekruksavanich, S., Jantawong, P. and Jitpattanakul, A. (2022), Resnet-based deep neural network using transfer learning for animal activity recognition, *in* '2022 6th International Conference on Information Technology (InCIT)'. Nonthaburi, Thailand, 10-11 November 2022, pp. 445-449. https://doi.org/10.1109/InCIT56086.2022.10067405.

Nature Conservation Division, Department of Forests and Park Services (2022), *Conservation Status of Red Panda (Ailurus fulgens) in South-Western Bhutan*, Nature Conservation Division, Department of Forests and Park Services. Available at: https://redpandanetwork.org/files/galleries/Conservation_Status_of_Red_panda_Final.pdf [Accessed 20 September 2024].

Neeli, S., Guruguri, C., Kammara, A., Annepu, V., Bagadi, K. and Ch, V. (2023), Bird species detection using cnn and efficientnet-b0, *in* '2023 International Conference on Next Generation Electronics (NEleX)'. Vellore, India, 14-16 December 2023, pp. 1-6. https://doi.org/10.1109/NEleX59773.2023.10420966.

Oion, M. S. R., Islam, M., Amir, F., Ali, M. E., Habib, M., Hossain, M. S. and Wadud, M. A. H. (2023), Marine animal classification using deep learning and convolutional neural networks (cnn), *in* '2023 26th International Conference on Computer and Information Technology (ICCIT)'. Cox's Bazar, Bangladesh, 13-15 December 2023, pp. 1-6. https://doi.org/10.1109/ICCIT60459.2023.10441585.

Priya, P. K., Vaishnavi, T., Selvakumar, N., Kalyan, G. R. and Reethika, A. (2023), An enhanced animal species classification and prediction engine using cnn, *in* '2023 2nd International Conference on Edge Computing and Applications (ICECAA)'. Namakkal, India, 19-21 July 2023, pp. 730-735. https://doi.org/10.1109/ICECAA58104.2023.10212299.

Red Panda Network (2024*a*), 'Findings from the first-ever red panda collar study'. Available at: https://redpandanetwork.org/post/Findings-from-the-First-Ever-Red-Panda-Collar-Study [Accessed 10 September 2024].

Red Panda Network (2024*b*), 'Threats to red pandas'. Available at: https://redpandanetwork.org/threats/ [Accessed 10 September 2024].

Sharma, D., Singh, J., Isha, Vipin, Panwar, D., D, A. and Anuj (2024), Animal classification using cnn architectures, *in* '2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)', Vol. 1. Greater Noida, India, 29-31 August 2024, pp. 1-6. https://doi.org/10.1109/ICEECT61758.2024.10738913.

Song, Z., Gong, W., Li, C. and Toe, T. T. (2022), Animals image classification method based on improved convolutional neural network, *in* '2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)', Vol. 10. Chongqing, China, 17-19 June 2022, pp. 1997-2001. https://doi.org/10.1109/ITAIC54216.2022.9836676.

Spiezio, C., Altamura, M., Weerman, J. and Regaiolli, B. (2022), 'Behaviour of zoo-housed red pandas (ailurus fulgens): A case-study testing the behavioural variety index', *Journal of Zoological and Botanical Gardens* **3**(2), 223–237. https://www.mdpi.com/2673-5636/3/2/18 [Accessed 25 September 2024].

Syed Zaini, S. Z., Sofia, N. N., Marzuki, M., Abdullah, M. F., Ahmad, K. A., Isa, I. S. and Sulaiman, S. N. (2019), Image quality assessment for image segmentation algorithms: Qualitative and quantitative analyses, *in* '2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)', pp. 66–71. https://doi.org/10.1109/ICCSCE47578.2019.9068561.

Valarmathi, B., Gupta, N. S., Prakash, G., Reddy, R. H., Saravanan, S. and Shanmugasundaram, P. (2023), 'Hybrid deep learning algorithms for dog breed identification—a comparative analysis', *IEEE Access* **11**, 77228–77239. https://doi.org/10.1109/ACCESS.2023.3297440.

Zhang, D., Liu, Z. and Shi, X. (2020), Transfer learning on efficientnet for remote sensing image classification, *in* '2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)'. Harbin, China, 25-27 December 2020, pp. 2255-2258. https://doi.org/10.1109/ICMCCE51767.2020.00489.