Image classification of birds species using deep learning approach

line 1: Muhammad Rizwan Tahir (1023340)  
line 2: *Liverpool John moores university*line 3: Liverpool, England  
line 4: rizwantahir730@gmail.com

*Abstract*— Scientists have begun collaborating in recent years to address the loss of animal species. Because it keeps the natural system stable, wildlife protection is crucial. Until now, many machine and deep learning approaches have been developed to recognize and classify different animal species. In this study, a task has been assigned by the BBC Autumwatch to investigate the use of deep learning approaches for the classification and identification of bird species with the dataset in the form of images. Different strategies have been used in this study's exploratory data analysis and data pre-processing on the dataset of birds so our data can be ready for model training. To classify the bird species, this study also proposes the use of three different architectures SSD (Single Short Detector), ResNet and EfficientNet. In future work, TensorFlow library will be used to implement all proposed three architectures, and the results will be compared to determine which model is the best.

Keywords—Artificial Intelligence, Machine learning, Deep learning, Transfer learning, BBC Autumwatch, SSD, ResNet, EfficientNet, wildlife, neural network, CNN

# Introduction

Maintaining the balance of the environment depends on wildlife protection. A number of natural processes in nature are also stabilised by it. The equilibrium of the ecosystem is greatly influenced by animals. In addition, because of the animals that inhabit there, forests are more visually appealing, which attracts people. Unfortunately, numerous species are in risk of extinction in the current year [1]. Climate change worldwide and ecological degradation are the main factors contributing to species extinction. According to [2],  identification of wildlife  is crucial for keeping track of the planet's declining biodiversity. As per [3], the environment of an endangered species must be continuously observed for changes in behaviour as well as current knowledge of the existence, location, and those changes.

For the wildlife management and conservation, automation is needed to recognise and categorise individuals and objects in visual fields [4]. There are numerous machine learning and deep-learning algorithms available today to recognise different types of wildlife species. Nevertheless, each approach has benefits and drawbacks that vary based on the kind of problem that has been solved.

The authors of prior studies used numerous ways to identify animal species using both conventional machine and advance deep learning approaches. The identification of bird species is the main emphasis of this research topic with the discussion of both conventional machine and advance deep learning approaches. In terms of identification of bird’s species, the author of [5], did research to identify bird species using a conventional method by employing brief DNA barcode patterns from a specific region of the genome. However, as evidenced by DNA sequence data and other molecular techniques that can lead to the hybridization of some species [6], invasion of the genome is frequent, which restricts the barcode classification. In [7] study focuses on automatically identifying bird species from their audio-recorded songs. Results of this research were obtained with MLP (Multi-Layer Perceptron) and SMO (Sequential Minimal Optimization) classifiers. However, background noise was the biggest limitation, author stated that, the audio files were collected from real surroundings, with no filtering or other pre-processing, and as a result, they received noise from the environment. Therefore, using a variety of traditional machine learning methods to identify distinct bird species has several drawbacks. But recently, deep learning methods have been improved, and the outcomes for recognising animal species using pictures have been remarkable [3]. Moreover, utilizing manual and traditional statistical methods to analyse the enormous amount of data has become costly, time-consuming, and labour-intensive. The deep learning research has recently made significant strides, and these advances are showing promise in automating the study of the enormous datasets [8]. Several deep learning-based architectures have been implemented so far, including VGG, MobileNet, SSD (Single Short Detector), ResNet and YOLO, with different convolutional layers of deep learning neural network, for the object detection of animal species [9].

For the object detection of bird species, this study proposes three architectures, SSD, Faster R-CNN paired with the ResNet feature map and EfficientNet for getting the highest performance and reaching the highest average mean precision. Each of these architectural designs has drawbacks and advantages. Results of these architectures will be compared to see which yields the best classification of bird species.

The remaining sections of this research are organised as follows: The challenge outlined by BBC Autumwatch, computer vision and bird’s species identification using deep learning will be covered at section II. The methodological part, which provides an explanation of the bird dataset and model architectures for object detection in section III, is followed by a discussion of the challenge and justification of proposed system in section IV, and the final section includes information on future work and an evaluation strategy.

# Background

Every species present in a habitat makes a substantial contribution to keeping it healthy. Birds and other native species are crucial for preserving the equilibrium of an environment. In the greater web of life, they are viewed as species that support steady population levels among all living things[10]. Even a small decrease in their population threatens[11] Birds could be found at several various levels throughout the trophic system. They are both mid-level consumers and top predators. By means of pollination or seed dispersal, they help plants return to their environment. Many birds are scavengers, removing dead animals swiftly and recycling nutrients to keep the ecosystem healthy. However, changes in avian distribution are frequently the initial indication of an environmental disturbance [12]. According to a survey by BirdLife International, many bird populations worldwide are in decline. There are approximately 1500 bird species that are close to going extinct, and it is difficult to predict when this will happen [11]. In order to conserve biodiversity, it is crucial to create a quick and effective system for identifying its components and monitoring changes through time [12].

Technology has a key role to play in biodiversity and ecological protection, and it may significantly reduce the time and effort required to complete the necessary activities [13]. BBC Autumwatch developed a tool for detection of British garden birds from the images that have been uploaded, together with an explanation of how and why the computer giving a certain prediction. During the development of the tool, they used pre-built ResNet34 architecture having 34 layers of CNN (Convolutional Neural Network), trained on ImageNet database containing 14 million annotated images. They used unbalanced dataset of iNaturalist as that was the dataset that offered the most training images for British bird species. While unbalanced dataset can ignore minority classes and can be biased towards the majority classes, they experimented with different re-sampling techniques such as, under-sampling and oversampling, and data augmentation. They tested under sampling and oversampling but found drawbacks of each of the techniques. It has been stated that under sampling can waste useful information and oversampling can affect the model to overfit. So, they found Synthetic Minority Over-sampling Technique (SMOTE), a method of data augmentation best technique and selected for their classification model [14]. Moreover, the challenges listed by BBC while creating a reliable object detection model include occlusion, changing lighting conditions, a wide range of scale, changing viewpoints, etc[15].

In prior studies, many researchers used image classification techniques of deep learning in order to detect or classify the bird’s species. In [16], researchers used DCNN (Deep convolutional neural networks) for the identification purposes. The author stated that they used TensorFlow library to implement DCNN and they used database of Caltech-UCSD Birds 200. The author of this study claims 89% accuracy from their experiments. However, the percentage of accuracy is not satisfactory. In [17], the VGG-16 neural network was utilized by the study's author to extract features from photos of birds. They used a dataset containing photographs of various bird species found in Bangladesh to accomplish the classification. These photographs were used without any annotation. Then, they employed a variety of classification techniques, each of which produced a unique set of results. The author stated that VGG-16 network provided the highest accuracy of 89% when compared to other categorization methods like Random Forest and K-Nearest Neighbor (KNN). However, the author compared the results with machine learning approaches not deep learning approaches. In [18], this study proposes the transfer learning method and used inception-ResNet-v2 for the identification and classification of bird species in Taiwan. The author of this study claimed that, through their experiments, their model was 100% accurate in identifying birds among various object categories and 98.39% accurate in classifying the bird species. To achieve effective feature selection of birds, they investigated numerous backbone models of various deep learning architectures, including MobileNetV2, ResNet101, Xception, InceptionV3 and Inception-ResNet v2. They chose Inception-ResNet-v2 as the deep learning architecture since it performs significantly better than other backbone models.

An earlier study found that experts have been investigating various technologies to carry out automatic species identification for years.

# Methodolgy

This section of the paper is divided into six parts. The data description of the dataset of birds will be in the first part, analyze it by using some graphs in the second part, and pre-process data in the third part, so the dataset can be modelled after being analyzed. A discussion of the selection of the model architecture will take place in the fourth part, a discussion of the model training will take place in the fifth part, and a final discussion of the inference and evaluation of the selected model will take place in the last part.

## Data Description

Monitoring bird diversity is a crucial undertaking from an environmental and ecological perspective. Despite being a well-established technique, bird monitoring has limited scalability because the observation is usually done manually and takes a lot of time. This has inspired the use of crowdsourcing, camera-trap data, recorded data, and machine learning techniques to evaluate bird images and noises.

To develop a deep learning object detection model that can be used to track the numerous bird species that visit gardens in the fall is the objective of the BBC Autumnwatch computer vision challenge. A standard camera trap that generates still image data with a resolution of 1024 x 768 pixels is the apparatus they intend to use.

The dataset of 4000 images has been provided by BBC Autumnwatch for the object detection model. There are 4 classes of bird’s species, each class contain 1000 images. The names of each class are listed below:

* Erithacus\_rubecula
* Periparus\_ater
* Pica\_pica
* Turdus\_merula

As, each class is distributed equally we can say that this dataset is balanced. However, when will perform image tagging for the object detection there is a possibility that we might lose some of the images from each class.

## Exploratory Data Analysis

Exploratory data analysis (EDA), that involves exploring and summarising datasets, is typically carried out by data scientists. EDA is used to identify glaring errors, gain a better understanding of data patterns, identify outliers, and discover correlation between variables. On the dataset for the bird, various visualisation approaches have been used to draw conclusions from the data. The insights of the images of bird’s dataset have been examined jupyter notebook. At first, we will import directory of dataset of images in python file to get insights and then will look at images in detail so we can perform operations accordingly.

**Raw Images of each class**

We can start out by just taking a quick glance at a few randomly selected images.

Graphical user interface, website

Description automatically generated

**Fig. 1 – Random Images of each class**

In Fig 1, we are plotting 3 images of each class randomly to make comparison between the raw images. This figure shows that each bird specie has its own characteristics so that’s how they are different from each other.

**Class Distribution**

Graphical user interface, text, application, email

Description automatically generated

**Fig. 2 – Class distribution of each class**

In the dataset given by BBC Autumwatch, there are four classes, each class represent each bird species. From the Fig 2, all the classes are equally distributed with 1000 images each, shows class balance. However, when will be performing image tagging or labelling for object detection it might be possible that some of the images in each class may be useless and need to drop or remove them. This will result in class imbalance, to make our classes balanced again we will use data augmentation techniques.

**Shape of Images**

Graphical user interface

Description automatically generated

**Fig. 3 – Shape of single image of two different classes**

Moving forward, let’s check the image resolution of two different classes. Fig 3 shows that resolution of the images is not the same in the dataset. We can see that the resolution of class “Periparus\_ater” is 580 x 435 while “Pica\_pica” has 1024 x 768 resolution. The purpose of checking the image resolution is to get the same size images for training. Although, same image resolution is not the requirement of CNN (Convolutional Neural Network) as it can handle this by applying the filters.

|  |  |
| --- | --- |
| Chart, scatter chart  Description automatically generated  **Erithacus\_rubecula** | Chart, scatter chart  Description automatically generated  **Periparus\_ater** |
| Chart, scatter chart  Description automatically generated  **Pica\_pica** | Chart, scatter chart  Description automatically generated  **Turdus\_merula** |

**Fig. 4 Dimension of images of each class**

Now let’s see the shape of every image in all 4 classes. In the table of Fig 4, jointplot() function of seaborn library shows the dimension of each image in each class. From the table, most of the images have different pixel values which can affect model training phase. In order to solve this problem, we can re-scale those images to the same size to reduce biasing of the system.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Class Name** | **Mean** | | **Standard Deviation** | |
|  | **Dim 1** | **Dim 2** | **Dim 1** | **Dim 2** |
| **Erithacus\_rubecula** | 802.727 | 932.021 | 160.431 | 155.264 |
| **Periparus\_ater** | 747.664 | 960.904 | 141.445 | 142.33 |
| **Pica\_pica** | 815.301 | 927.924 | 158.185 | 150.077 |
| **Turdus\_merula** | 847.993 | 888.05 | 173.365 | 165.663 |

**Fig. 5 Mean and standard deviation**

Table in Fig 5 represents the mean and standard deviation values of each class. In class ‘Erithacus\_rubecula’, mean value of the width of all the images is 802.727 and the mean value of height is 932.02. While standard deviation of each dimension in this class is 160.431 and 155.264. For the normalization of images, the mean value must be 0 and standard deviation must be equal to 1. However, in this dataset we can see high standard deviation, which shows that data is not normalised. Data normalization is important step in data pre-processing of neural networks to improve learning rates.

## Data pre-processing

Pre-processing data is a typical second step in the deep learning workflow following data analysis in order to transform raw data into a format that the network can understand. Image tagging or annotation for the object detection, Data augmentation, Normalisation and dimension reduction can be the processes of data pre-processing.

**Image Tagging or Annotation**

**A picture containing bird, outdoor, oscine, standing

Description automatically generated**

**Fig. 6 – Image tagging for all classes**

Image tagging has been done for all the 4000 images in the dataset. The purpose of doing this is to identify object classes inside the image. Essentially, image tagging makes the model easy to understand the content of the image. Moreover, annotations and labelling of images let computers recognise objects more accurately, which enhances computer vision.

Fig 6 shows the tagged image of each class. This process has been done using the ‘ReNomTAG’ tool. After image tagging, we have found that some of the images in the dataset were not containing any object related to birds species so we discard them. From the total of 4000 images, 3879 were tagged and 121 images were dropped out.

**Data Augmentation**

During the image tagging process we found some images which were not useful, so we removed them from the dataset, results in class imbalance. To handle class imbalance there are several techniques such as re-sampling. However, as mentioned by [14]the data augmentation is the best technique for the balancing. Consequently, we are utilising this method for our classification model.

The method of TensorFlow ‘ImageDataGenerator’ has been used to generate different images that our dataset does not have.

A picture containing text, bird, oscine, perched

Description automatically generated

**Fig. 7 – Newly generated image from existing one**

Fig 7 shows the output of image generator method of TensorFlow.

**Normalisation**

As discussed before, from Fig 4 and Fig 5, our data is not normalised. Therefore, normalisation is a pre-processing technique for standardising data. Having several data sources with comparable ranges. If the data is not normalised before training, our network may encounter problems, making it far more difficult to train it and lowering its learning rate. As a result, in our future work, we will use the batch normalisation technique for our dataset.

**Dimensionality Reduction**

Finally, by utilising a dimension reduction approach like principal component analysis, we can see the elements that best characterise each class (PCA). It is possible to reorganise and display the eigen pictures of our image matrix, which are actually the eigenvectors (components) of PCA.

**Table

Description automatically generated**

**Fig. 8 – Dimensionality reduction of Erithacus\_rubecula**

Fig 8 shows that we used PCA to reduce the dimensionality of ‘Ertithacus\_rubecula’ class data. PCA has been applied to all the classes. The purpose of doing this is to interpretability of the model by finding the most important features in the data.

## Model Architecture Selection and Justification

For the object detection of diverse animal species, several deep learning-based architectures have been developed, including VGG, MobileNet, SSD (Single Short Detector), ResNet, and YOLO, using various convolutional layers of deep learning [9].

The dataset in the form of images has been provided by BBC Autumwatch for the object detection of bird species. The process of tweaking the filters in the convolutional layers when training CNNs (Convolutional Neural Networks) from scratch takes many images and processing resources. So, this study suggests transfer learning with pre-build architectures with convolutional layers. The SSD, Faster R-CNN with the ResNet feature map, and EfficientNet architectures are the three that this study selected for the implementation in future work as they have the best performance and average mean precision. Every one of these architectural concepts has benefits and drawbacks. Like SSD cannot perform well for the small objects in the image whereas the biggest limitation with Fast R-CNN is its speed. To determine which architecture produces the best classification of bird species, the results of different buildings will be compared.

## Model Training

The dataset will be divided into training and test sets for model training. The training set will have 80% of the data and 20% will be in the test set. After splitting the data into different sets, we will set hyperparameters for the method we will be using. The purpose of hyperparameters is to control the learning process of the model. Let’s discuss the model training of all the three selected architectures with the best hyperparameters.

**SSD Architecture**

An SSD is made up of two components: a backbone model and an SSD head. In most situations, the feature extractor of the backbone model is a pre-trained image classification network. In this case, networks trained on ImageNet, like MobileNet, have typically been utilized, however without the final fully linked classification layer. The only convolutional layers that are added to this backbone by the SSD head are one or more, and their outputs are read as object classifications and bounding boxes in the areas where the final layer's activations occurred [19].

The hypermeters that will be used for the training of this model are training epochs, batch size, optimizer, learning rate, momentum and decay.

**ResNet Architecture**

ResNet is a pretrained network, can classify images into many object categories and it can be any category such as laptop, cup, phone and many living things. ResNet can go to several layers deep, for example, ResNet-50 is 50 layers deep and ResNet-152 is 152 layers deep.

The hypermeters that will be used for the training of this model are training epochs, batch size, optimizer, initial learning rate, learning rate decay factor, end learning rate, momentum and weight decay rate.

**EfficientNet Architecture**

EfficientNet is currently the most successful convolutional neural network for classification. EfficientNet employs a technique known as compound coefficient to quickly and simply scale up models. Compound scaling use a specified set of scaling factors to consistently scale each dimension as opposed to employing a random scaling strategy.

The hypermeters that will be used for the training of this model are training epochs, batch size, optimizer, learning rate momentum and weight decay rate.

## Inference and Evaluation

The next step after the model has been trained is to infer from and assess the trained model. The model will then be used to process external data, enabling it to find and recognize anything that it has been trained to recognize. It's not good practice to train your model on the entire dataset. Divide the data into training and test sets. It is best to split the dataset into training and test sets at 70–80% and 20–30%, respectively. Furthermore, it is essential to include new data while analyzing it in order to prevent the potential of models getting overfit to the training set. Moreover, the model must not be underfitted, which shows a significant degree of bias and ignores the external data.

Several metrics, including the bias-variance trade-off, confusion matrix, accuracy, precision, recall, and mean squared error, will be used to assess the model.

# Discussion

Animal preservation is essential for maintaining the ecological balance. However, a number of species are sadly in danger of being extinct this year [1]. Numerous bird species are become harder to find, and even when they are, it could be challenging to predict their categorization. Birds can be seen in a variety of sizes, shapes, colours, and orientations when viewed up close. The breed of the bird is much more variable in the pictures than it is in the auditory classification. People are better able to distinguish between different birds when they use images [16]. In Artificial Intelligence, scientists are working intensively to detect and preserve various animal species. Regression, classification, clustering, and other types of problems can all be solved using a variety of pre-built machine learning and deep learning algorithms.

In this paper, the dataset of birds in the form of images has been used to perform EDA (Exploratory Data Analysis) and Data pre-processing techniques so the data is prepared for the model training. SDD, ResNet and EfficientNet, three different architectures of deep learning have been proposed for the object detection of bird’s species. Prior studies show the effectiveness of these architectures, which yields good results, among proposed architectures EfficientNet is the best one performance wise. SSD requires less computational requirements but cannot handle small objects on the other hand ResNet can deal with the small objects in the image. Each architecture has its own benefits and drawbacks depending on the problem. In the future work all of these will be implemented and then results of these architectures will be compare between each other to get to know that which architecture gives the best results.

# Future work

In the future work, all the proposed architectures will be implemented using TensorFlow library. Transfer learning approach will be used for the implementation of these architectures. All the proposed architectures, SSD, ResNet and EfficientNet, will be compared with each other to check which architecture yields the best results. For the evaluation of these architectures different evaluation metrics will be used such as bias-variance trade off, confusion matrix, accuracy, precision, recall, and mean squared error. The evaluation of the models will also be done by external test data.

##### References

[1] K. M. Iraqi, IEEE Computer Society, University of Karachi. Department of Computer Science, and Institute of Electrical and Electronics Engineers, *ICISCT’20 : 2nd International Conference on Information Science and Communication Technology : 8th-9th February 2020*.

[2] C. Duhart, G. Dublon, B. Mayton, G. Davenport, and J. A. Paradiso, “Deep Learning for Wildlife Conservation and Restoration Efforts.”

[3] S. B. Islam and D. Valles, “Identification of Wild Species in Texas from Camera-trap Images using Deep Neural Network for Conservation Monitoring,” in *2020 10th Annual Computing and Communication Workshop and Conference, CCWC 2020*, Jan. 2020, pp. 537–542. doi: 10.1109/CCWC47524.2020.9031190.

[4] Z. Miao *et al.*, “Insights and approaches using deep learning to classify wildlife,” *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-019-44565-w.

[5] P. D. N. Hebert, M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis, “Identification of birds through DNA barcodes,” *PLoS Biol*, vol. 2, no. 10, Oct. 2004, doi: 10.1371/journal.pbio.0020312.

[6] J. Mallet, “Hybridization as an invasion of the genome,” *Trends in Ecology and Evolution*, vol. 20, no. 5. Elsevier Ltd, pp. 229–237, 2005. doi: 10.1016/j.tree.2005.02.010.

[7] M. T. Lopes, L. L. Gioppo, T. T. Higushi, C. A. A. Kaestner, C. N. Silla, and A. L. Koerich, “Automatic bird species identification for large number of species,” in *Proceedings - 2011 IEEE InternationalSymposium on Multimedia, ISM 2011*, 2011, pp. 117–122. doi: 10.1109/ISM.2011.27.

[8] A. Mirugwe, J. Nyirenda, and E. Dufourq, “EPiC Series in Computing Automating Bird Detection Based on Webcam Captured Images using Deep Learning,” 2022. [Online]. Available: https://news.mongabay.com/2017

[9] T. M. T. , C. Y. , S. W. , P. S. , M. S. N. , B. C. V. E. , A. A. S. A. , and V. K. A. Md Zahangir Alom1, “Deep\_learning\_approaches”.

[10] T. Telenský, P. Klvaňa, M. Jelínek, J. Cepák, and J. Reif, “The influence of climate variability on demographic rates of avian Afro-palearctic migrants,” *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-74658-w.

[11] BirdLife International, “Birds - BirdLife International”, Accessed: Nov. 23, 2022. [Online]. Available: https://www.birdlife.org/birds/

[12] Y. P. Huang and H. Basanta, “Recognition of Endemic Bird Species Using Deep Learning Models,” *IEEE Access*, vol. 9, pp. 102975–102984, 2021, doi: 10.1109/ACCESS.2021.3098532.

[13] P. C. Ravoor and S. T.s.b., “Deep Learning Methods for Multi-Species Animal Re-identification and Tracking a Survey,” *Computer Science Review*, vol. 38. Elsevier Ireland Ltd, Nov. 01, 2020. doi: 10.1016/j.cosrev.2020.100289.

[14] ben Hughes, “A machine’s guide to birdwatching - BBC R&D”, Accessed: Dec. 03, 2022. [Online]. Available: https://www.bbc.co.uk/rd/blog/2021-05-a-machines-guide-to-birdwatching

[15] J. U. Kim, J. Kwon, H. G. Kim, and Y. M. Ro, “BBC Net: Bounding-Box Critic Network for Occlusion-Robust Object Detection,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 4, pp. 1037–1050, Apr. 2020, doi: 10.1109/TCSVT.2019.2900709.

[16] P. Anusha and K. Manisai, “Bird Species Classification Using Deep Learning,” in *2022 International Conference on Intelligent Controller and Computing for Smart Power, ICICCSP 2022*, 2022. doi: 10.1109/ICICCSP53532.2022.9862344.

[17] S. Islam, S. I. A. Khan, M. Minhazul Abedin, K. M. Habibullah, and A. K. Das, “Bird species classification from an image using VGG-16 network,” in *ACM International Conference Proceeding Series*, Jul. 2019, pp. 38–42. doi: 10.1145/3348445.3348480.

[18] Y. P. Huang and H. Basanta, “Recognition of Endemic Bird Species Using Deep Learning Models,” *IEEE Access*, vol. 9, pp. 102975–102984, 2021, doi: 10.1109/ACCESS.2021.3098532.

[19] C. Tang *et al.*, “Object detection method of multi-view SSD based on deep learning,” *Hongwai yu Jiguang Gongcheng/Infrared and Laser Engineering*, vol. 47, no. 1, Jan. 2018, doi: 10.3788/IRLA201847.0126003.