Bird Image Classification using Convolutional Neural Network Transfer Learning Architectures

Asmita Manna[1], Nilam Upasani[2], Shubham Jadhav[3], Ruturaj Mane[4], Rutuja Chaudhari[5], Vishal Chatre[6]

Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India[1, 3, 4, 5, 6]
Balaji Institute of Technology and Management, Sri Balaji University, Pune, India[2]

Abstract—**With the technological progress of human beings, more and more animal and bird species are being endangered and sometimes even going to the verge of extinction. However, the existence of birds is highly beneficial for human civilization as birds help in pollination, destroying harmful insects for crops, etc. To ensure the healthy co-existence of all species along with human beings, almost all advanced countries have taken up some conservation measures for endangered species. To ensure conservation, the first step is to identify the species of birds found in different locations. Deep learning-based techniques are best suited for the automated identification of bird species from the captured images. In this paper, a Convolutional Neural Network based bird image identification methodology has been proposed. Four different transfer learning-based architectures, namely Resnet152V2, Inception V3, Densenet201, and MobileNetV2 have been used for bird image classification and identification. The models have been trained using 58388 images belonging to 400 species of birds, and the models have been tested using 2000 images belonging to 400 species of birds. Out of these four models, Resnet152V2 and DenseNet201 performed comparatively well. The accuracy of Resnet152V2 was highest at 95.45%, but it faced a large loss of 0.8835. But based on the results, even though DenseNet201 had an accuracy of 95.05%, it faced less loss i.e., of 0.6854. The results show that the DenseNet201 model can further be used for real-life bird image classification.**

Keywords—Deep learning; CNN; Image classification; DenseNet201; ResNet152V2; InceptionV3; MobileNetV2

I. INTRODUCTION

Biodiversity around us is quite important for human civilization as it helps maintain a balance in the ecosystems. Birds are one of the most important resources, which help us in various ways like pollination, protecting crops from harmful pests destroying the crops, etc. However, with rapid industrialization, more and more bird species are becoming endangered and sometimes on the verge of extinction. Bringing more forestland and wetland under modernization is also fueling the extinction process. Therefore, almost all countries have taken one or more conservation projects to preserve biodiversity and conserve endangered species [1][2].

To preserve a species, the first important step is to identify the species correctly followed by the appropriate steps to preserve it. Deep learning-based approaches are best suited for the automated identification of bird species because deep learning enables us to extract the features of birds and provide higher accuracy while predicting. Deep learning can identify the species of a bird based on data in the form of audio, video, and image. In the case of real life, audio and video are not that suitable as the overlapping of chirping of multiple birds can't be ruled out, and those noises may tamper with the essential features of the bird. Thus, in this paper, a deep learning-based approach trained with bird images has been proposed. There are some papers which proposed the use of images for bird species identification [9][20][21]. But the novelty of our study is that we have included 400 species for identification which has expanded the application of our study.

As there are hundreds of deep learning algorithms available, it is necessary to investigate with the test data set to understand which algorithm generates the best accuracy and minimizes the loss. In this paper, four different transfer learning-based models namely Resnet152v2, InceptionV3, Densenet201, and MobileNetV2 are tested and compared for bird image classification. The transfer learning models are used to extract the features of the birds. The dataset used contains a total of 58388 images of birds. The total number of species included in the dataset is 400 [31]. The output and input layer are added to the transfer learning model with dense connection. The output layer is generating results by using SoftMax. Given that SoftMax is employed for multiclass classification, it serves as the activation function.

The major research objectives of this paper are as follows: (a) To compare various transfer learning models for bird species identification from bird image datasets. (b) To build a system of automatic identification of bird species with maximum accuracy using the transfer learning model which is most suited for the given dataset. (c) To put forth a web-based system which will help photographers to build their portfolio by identifying the bird species from the image they captured.

A. Contribution of the Paper

- A comprehensive study of existing deep learning/ transfer learning-based systems for bird species identification and pinpointing the drawbacks of those existing systems.

- Identification of an intricate dataset for bird species identification by studying various related datasets.

- Implementation of different transfer leaning models such as InceptionV3, ResNet152V2, DenseNet201 and MobileNetV2 on the identified dataset.

- Selection of the most efficient and accurate model suited for image classification of birds and the measures that can be incorporated to increase the performance.

Section II of this paper presents the literature review to find the gap in the existing approaches; Section III presents the proposed architecture and methodology; Section IV shows the experimental results and, finally Section V concludes the paper.

II. LITERATURE REVIEW

Many researchers have used Deep Learning and Convolutional Neural Network based approaches for image classifications, especially for identifying diseases from image datasets, but not much work is done for identifying bird species from the image datasets of birds. In this section, the existing approaches of image classification using transfer learning techniques are explored.

A recent paper [3] explaining the deep learning architecture for the detection of birds from the images captured by the webcam was published. For the pre-processing of data, deep learning algorithm's capability for the detection of birds inside the images is checked first. Then the authors used two CNN models i.e., single-shot detectors (SSD) and Quicker R-CNN as a combination with Inception ResNet152V2, ResNet50, ResNet152, ResNet101, and MobileNet-V2 features. By combining a faster R-CNN, they got high precision, and the SSD with MobileNetV2 was selected as the best model in terms of speed and smaller memory consumption.

Transfer learning model was used with six different CNN architectures such as DenseNet201, MobileNetV2, ResNet50, InceptionResNetV2, ResNet152V2, and Exception in a paper [4]. After calculating the evaluation matrix, the authors found that MobileNetV2 performed better than the other transfer learning approaches in terms of evaluation matrices. However, they could classify the bird images in only eight categories.

A paper was published in which [5] the authors worked to find a suitable model for transfer learning. After comparison of results, the authors showed that the training Inception-v3 on the CIFAR-10 dataset provided better results. The authors also explained that the basic to advanced use of transfer learning may be used not only for the model presented but also for other deep neural networks for image classification.

A paper [6] implementing the CNN algorithm to extract information from bird images was published. The CNN model was developed entirely from scratch. The model was then trained to test its efficiency. The developed application had a high accuracy of about 93.19% on the training set and 84.9% on the testing set.

The image classification models MobileNetV2 and Inception-v3 were proposed to be used in a paper [7]. The authors used four approaches—Inception-v3 with and without transfer learning and MobileNetV2 with and without transfer learning—to accomplish the task. Among the four approaches, MobileNetV2 with transfer learning performed better, with an accuracy of about 91.00%.

In a paper published in 2021[8], the authors tried to identify the habitat elements from the bird elements using a deep convolutional neural network and ResNet152 dependent models, which gave the test accuracy that was best. It has been proven that a deep convolutional neural network could be effective for automatically identifying habitat elements from images of birds. The author concluded that the actual implementation of this technology would be extremely useful in understanding the relationship between habitat elements and birds.

The paper [9] proposed a VGG-16 network model-based solution to extract bird image features. The authors used different classification methods, each with different results. Support Vector Machine (SVM), when compared to other categorization techniques like random forest and K-Nearest Neighbor (KNN), provided the highest accuracy of 89%.

The authors of a paper which was published in 2019 [10] developed a cloud-based mobile phone app that makes use of deep learning for image processing to find the species of birds from the various digital images that are transferred by the user over a mobile phone. Convolutional Neural Network was trained using bird images to limit the outstanding image features. The Convolutional Neural Network model with bound connections gave high accuracy, which is 99.00%, as compared to the CNN with 93.98% and the SVM with 89.00% for the image training.

A pose-normalized deep convolutional nets approach was proposed in a paper published [11]. The proposed method depends on a detection part, and the Convolutional Neural Network features are extracted from various regions that are pose-normalized. Its execution was better for the usage of Convolutional Neural Network features, which were fine-tuned upon CUB-200-2011 for every region. It also was effective for the usage of various Convolutional Neural Network layers for numerous alignment levels; and the usage of a similar warping function that's estimated to use a larger number for detecting key points. They introduced an innovative method for studying a group of various pose-regions that clearly curtails the alignment of pixel error. Then it works on complicated pose-warping operations.

Another paper [12] proposed a deep learning model which can identify individual birds from the input image. The authors proposed two different models and showed that the proposed pre-trained ResNet model has achieved better accuracy than the based model. The best model showed 97.98% accuracy in identifying bird species.

The paper proposed in 2020 [13] used deep learning models to detect pneumonia based on chest x-ray images. The author used four models: two pretrained models (ResNet152V2 and MobileNetV2), CNN, and long-short-term memory (LSTM). The result showed that the ResNet152V2 performed better and the MobileNetV2, CNN, LSTM accuracy, recall, F1-score, precision, and AUC were higher than 91%.

An author published a paper based on declining the North American avifauna [14] represented that by using the multiple and independent monitoring networks author represented the loss of population of the North American avifauna which includes the common species.

The paper proposed in 2021 describes the declining of the forest bird species [15] where the author studied birds in six

land types in the oak forest biome of Himalaya. The richness of the species was west in pine and built-up sites as compared to natural oak. The forest specialist and insectivores are reduced up to 60 - 80% in the modified forests.

The loss found in biodiversity in the European Union is described in a paper [16] by the author by studying the extensive dataset. The author evaluated that there is a decline of around 17-19 % in the overall bird abundance. The authors are suggesting that we should preserve the bird species and that it is beneficial to nature and human beings.

In the paper that was published in 2021 [17], the authors tried to explain that there is a huge amount of decrease in the bird species, some have vanished, some are endangered. As a result, it has a negative impact on both biodiversity and human lives. So, it's our responsibility to preserve them.

A paper that uses a deep convolutional neural network [18] to identify birds' images was published. The author used habitat elements of bird images. Author used the model based on the ResNet152 algorithm, given 95.52% validation and AlexNet given the lowest test accuracy as 89.48%. It is proved that a deep convolutional neural network is efficient and useful for bird image classification.

A paper was published which explained that ecological resources are important for the survival of human beings [19]. The author has used ecological research. The author first developed the relationship between the theory of deep learning and ecological research. It is expected that participation and preparing cross-disciplinary abilities may advance standardization. Deep learning is used for nonlinear feature extraction for scientific and industrial data processing.

A paper that was published in 2019 [20], used the VGG 16 network to extract features of birds. The author used a dataset of bird species of Bangladesh. The author used different classification methods, like random forest, K-nearest neighbor (KNN) but the support vector machine (SVM) gave the max accuracy of 89%.

A paper that uses a deep learning platform to identify bird species images [21] using the mobile app Internet of birds. The author used convolutional neural network (CNN) to find different features in images. To improve feature extraction, the skip connection method is used. Then the SoftMax function is used to get a probability distribution of the features of birds. 93.98 % was convolutional neural network (CNN) accuracy and support vector machine (SVM) got 89.00% accuracy. Both accuracies are less than 99.00%, which is the highest accuracy of the proposed model convolutional neural network (CNN) with skip connection.

The paper published in 2014[22] proposed architecture which first finds the pose of a bird. For bird feature extraction Deep Convolutional Neural network is used. To find the compact pose the author proposed a novel graph- based clustering algorithm. Author got great classification accuracy that is 75% vs. 55-65%.

A paper that was published in 2020 [23] proposed a deep learning model to identify bird species. The author used the ResNet model as a pre-trained convolutional neural network

(CNN) with a base model to identify the images. The author got a high accuracy of 97.98% on the bird classifications.

A paper [24] for predicting breast cancer was published in 2021 that used a dataset of QIN-Brest for breast cancer detection that is divided into the ratio 7: 3 for training and testing. The authors used two deep transfer learning models, DenseNet201, ResNet152v2 and an ensemble model with concatenation of two models, trained and tested using a dataset of CT images. The ensemble model has been given 100% accuracy on the test data. The authors concluded that ensemble models are better at predicting breast cancer than those of DenseNet201 and ResNet152v2 models.

A paper published in 2019 [25] explaining that there is a loss in abundance of bird species that leads to the changes in the ecosystem. Authors studied the bird population of North American avifauna for over 48 years. It is found that there is a loss in the population of birds, around 29% of 1970 bird population. This population loss needs to be addressed for the future of biodiversity.

In the paper that was published in 2021 [26], it is said that because of the land use change, there is a loss of biodiversity in different countries. That is responsible for changing the forest ecosystem. It is very important to know these things to avoid biodiversity loss. The author carried out a semantic breeding-season survey in six different land use types. The author's study shows that there is moderate to drastic species loss in all-the changed land uses compared to natural oak forests.

The Global Assessment Report on Biodiversity and Ecosystem Services is a thorough and evidence-based analysis of the state of biodiversity worldwide [27]. It pinpoints the causes of loss and the impacts on food security, health, and livelihoods. With recommendations for policy action, it provides a comprehensive understanding of the risks and opportunities associated with ecosystem degradation. Despite its strengths, future work is needed to address data gaps, regional limitations, implementation challenges, and the integration of other global challenges, requiring ongoing monitoring and evaluation.

The paper 'Abundance decline in the avifauna of the European Union' [28] reveals cross-continental similarities in biodiversity change" analyzes the changes in bird abundance in the European Union (EU) between 1980 and 2015. The study found that bird abundance has decreased by 22.5% in the EU during this period, and that the decline was most pronounced in farmland and grassland bird species. The decline in bird populations was found to be similar to trends observed in North America, suggesting cross-continental similarities in biodiversity change. The paper highlights the need for conservation efforts to reverse the decline in bird populations and prevent further biodiversity loss.

In a paper named 'Birds in Decline' [29], Youth H. takes us on a soaring journey through the troubling topic of declining bird populations. With a keen eye for detail and a passion for avian conservation, Youth H. paints a vivid picture of the challenges faced by birds and the ecosystems they inhabit. Through a mix of data analysis and personal

observations, the author sheds light on the alarming trend of bird decline and highlights the urgent need for action. The result is a thought-provoking and informative piece that will leave readers with a renewed appreciation for the feathered friends that share our planet.

The paper 'Applications for deep learning in ecology' [30] provides an overview of the use of deep learning in ecology. It explains how deep learning, a branch of machine learning, has become popular due to its flexibility and performance. The paper reviews existing implementations and demonstrates how deep learning has been used successfully to identify species, classify animal behavior, and estimate biodiversity in large datasets like camera-trap images, audio recordings, and videos. It also provides guidelines, recommendations, and useful resources to help ecologists get started with deep

learning. The authors argue that deep learning can become a powerful reference tool for ecologists, especially at a time when automatic monitoring of populations and ecosystems generates vast amounts of data that cannot be effectively processed by humans anymore.

The papers studied during the survey included bird image classification approaches, but the results were provided for only a few numbers of bird species. The dataset used for comparison of the transfer learning models in our study included 400 species of birds and the size was also large enough which ensured proper training and testing of the models to provide the maximum output from the models. A detailed comparative study of the discussed papers is presented in Table I.

TABLE I. COMPARATIVE STUDY OF EXISTING APPROACHES

Paper id	Dataset	Approach	Advantages	Limitations / Future work
[3]	Live bird feed watchers came from Cornell Lab of Ornithology at 6 different locations.	Single-shot detector and faster R-CNN with ResNet50, MobileNet-V2, ResNet152, ResNet101, and Inception ResNet-V2 feature extractors.	Faster R-CNN combined with ResNet152 feature extractor was best for achieving high precision.	The SSD with MobileNetV2 was best for fast speed and low memory consumption but lacked high precision like Faster R-CNN with ResNet152.
[4]	500 images of each bird class of 7 species of Bangladesh birds with a total of 2800 training images and 700 testing images	Transfer learning used in 6 different CNN architectures DenseNet201, InceptionResNetV2, MobileNetV2, ResNet50, ResNet152V2, and Xception.	MobileNetV2 outperforms all the other transfer learning models.	There is no support for using the live bird feed of the cam for identification.
[5]	CIFAR-10 and Caltech Faces dataset	Inception- v3 via transfer learning.	Transfer learning models are better than custom CNN.	Accuracy can be improved by increasing the epoch sizes and size of the dataset and combining CNN with Long-short Term Memory (LSTM).
[6]	Custom dataset using Microsoft's Bing Image Search API v7.	Used a deep learning model to extract the information images of birds using CNN.	Developed a smaller and more portable version of VGGNet and achieved an accuracy of 93.19% on the training dataset.	The testing dataset accuracy is 84.91% which is very low.
[7]	700 testing images and 2800 training images of 7 local birds in Bangladesh.	Used 4 approaches inception - v3 with and without transfer learning, MobileNet with and without transfer learning.	MobileNet performed better than the Inception model.	Only 7 bird species are used to generate the output.
[8]	Custom dataset obtained from BirdFans in China containing 20000 images.	Implemented proposed method with 4 types of deep CNN.	DCNN gave a high accuracy of 97.76% on the testing dataset.	Further research is needed to improve the DCNN for performing well in real-world scenarios.
[9]	1600 images of 27 species of birds of Bangladesh.	Used machine learning approach using VGG-16 network as a model.	SVM gave the maximum accuracy of 89%.	KNN has less accuracy than SVM.
[10]	27 Taiwan endemic bird species.	Given bird images as an input to CNN which is used with the skip connection.	The Convolutional Neural Network model with bound connection gives high accuracy which is 99.00 %.	CNN and SVM have less accuracy than CNN with the skip connection.
[11]	CUB-200-2011	The bird's features are computed by CNN.	Performance was better for 1] usage of Convolutional Neural Network features which were fine-tuned upon CUB-200-2011 for every region, 2] usage of various Convolutional Neural Network layers for numerous alignment levels type, and 3] usage of a similar warping function that's estimated for using a larger number for detecting key points.	For future work, the proposed solution can be applied to other fine-grained datasets. Then they explore the custom-built CNN network structures and their training.
[12]	Data of bird species from various sources merged with the western dataset.	Pretrained CNN performs better on input images.	The proposed pre-trained ResNet model has better accuracy of 97.98% in identifying bird species.	The based model showed less accuracy than the proposed pre-trained ResNet model.

[13]	QIN-Breast Dataset	Two deep transfer learning models, ResNet152V2 and DenseNet201, and an ensemble model with a combination of 2 models, trained and tested by CT images.	The proposed model gained the maximum accuracy of 100 percent on the dataset which was tested and also high performance of 100% in f1-score, recall and precision value.	Both the models still have scope of improvement as considering various parameters.
[14]	529 bird species from the US and Canada.	Multiple and independent monitoring networks (radar networks)	Networks used in the paper are useful for getting the population loss across the North American avifauna.	As the birds are giving numerous benefits to the ecosystem, people should conserve them.
[15]	8549 bird observations.	The landscape is chosen and further it is divided into parts as per their use.	There should be conservation planning of bird species in human dominated landscapes.	Due to the use of land by humans affected the forest bird species in western Himalaya.
[16]	445 native bird species.	Bayesian models are used to check the change in abundance of the birds in the EU. The imputed model shrinks the uncertain indices of the species towards the group mean.	Using the models used in a paper, the author represented the biodiversity loss in native avifauna.	There is a decline of 17-19% in the overall breeding bird abundance
[17]	Worldwide different bird species.	Different bird species were studied all over the world, and various vanished and endangered species were discovered.	Preserve the bird species.	There is a huge amount of decrease in the bird species, and it has a negative impact on both biodiversity and human lives.
[18]	Habitat Elements from Bird Images	Used habitat elements of bird image and model based on the ResNet - 152 algorithm, given 95.52% validation and AlexNet given the lowest test accuracy as 89.48%	ResNet basel model given highest validation accuracy is 95.52 %.	ResNet basel model given lowest test accuracy is 95.52 %.
[19]	Multi-source data	Used ecological resources that are important for the survival of human beings. Developed the relationship between deep learning and ecological research .Deep learning is used for nonlinear feature extraction	Understanding of ecological resources is beneficial for the sustainable development of human society.	-
[20]	Bird species of Bangladesh.	Used the VGG 16 network to extract features of birds. Different classification methods, like random forest, K-nearest neighbor(KNN) but the support vector machine(SVM) gave the max accuracy of 89%	Support Vector Machine (SVM) gave the max accuracy of 89%	Improve accuracy by increasing data
[21]	27 bird species endemic to Taiwan.	Mobile app is used to identify bird species images. Used CNN for features extraction. To improve feature extraction, the skip connection method is used.	99.00%, is the highest accuracy of the convolutional neural network (CNN)model with skip connection.	CNN has 93.98 %,SVM has 89.00% accuracy which is lesser than CNN with skip connection model.
[22]	Worldwide different bird species.	Finds the pose of a bird. Deep convolutional network is used for feature extraction. For compact poses a novel graph-based clustering algorithm is used.	Classification accuracy, is 75% vs 55-65% which is of old method.	-
[23]	Birds found in diverse scenarios worldwide.	Used resNet model as a pre-trained CNN network with a base model to identify the images.	High accuracy of 97.98%	Birds are found in diverse scenario with different size and shapes.
[24]	Public dataset QIN Breast	Used 2 deep transfer learning models, DenseNet201,ResNet152v2 and an ensemble model with concatenation of 2 models, trained and tested using a dataset of CT images.	Ensemble model has given 100% accuracy on the test data.	Ensemble models are better than DenseNet201 and ResNet152v2 models.
[25]	North American avifauna	Loss in abundance of bird species that leads to the changes in the ecosystem	Population loss is studied which is helpful for the future of biodiversity	American avifauna when studied it is found that a loss in the population of birds, around 29% of 1970 bird population.

[26]	Western Himalaya bird species.	Because of the land use change, there is a loss of biodiversity.	Studied land use change which will help in future to avoid biodiversity loss.	Because of the land use change there is decrease in the biodiversity.
[27]	-	It provides a comprehensive analysis that identifies the causes and impacts of biodiversity loss.	It provides policymakers with actionable recommendations, Integrated approach, Global perspective, Foundation for future research.	Areas for future work, such as data gaps, Limited regional focus and implementation challenges.
[28]	Extensive dataset of breeding bird abundance in the European Union.	The approach used in the paper was to analyze data from bird monitoring programs across the European Union to estimate population changes over time.	The authors estimated a decline of 17-19% in overall breeding bird abundance, equivalent to a loss of 560-620 million individual birds. The study highlights the high declines in bird numbers among species associated with agricultural land.	Limitations of the paper are that it did not include other aspects of biodiversity, only included bird populations in the European Union and did not consider populations in other regions, did not investigate the specific causes of bird population decline.
[29]	-	The paper examines senescence variation in multicellular organisms, particularly social species. It reviews senescence research, quantifies covariation between mortality and reproductive ageing, models social interaction effects on ageing, tests predictions in social species, and examines senescence in a cooperative breeder population.	This work can inform future research on the evolution of senescence and the dynamics of social species.	Future work for this paper could involve further testing and refinement of the social interaction model for senescence in social species.
[30]	-	The article discusses the potential applications of deep learning in ecology, including species identification, animal behavior classification, and biodiversity estimation. The authors provide guidelines, recommendations, and a reference flowchart for ecologists to get started with deep learning. They argue that as ecological datasets become larger and more complex, deep learning could become a powerful tool for ecologists.	It provides an overview of deep learning in ecology, demonstrates the usefulness of deep learning in ecology and highlights the potential of deep learning for automatic monitoring.	The paper does not provide an in-depth discussion of the limitations and potential biases of using deep learning in ecology.

III. PROPOSED WORK

The principal aim is to implement various transfer learning models to compare their performance on the bird image classification dataset. The main transfer learning models used for comparison are InceptionV3, ResNet152V2, Densenet201, and MobilenetV2. The transfer learning models are used as feature extractors. The dataset used contains a total of 58388 images of birds. The total number of species included in the dataset is 400. The output and input layer are added to the transfer learning model with dense connection. The output layer is generating results by using SoftMax as the activation function. Given that SoftMax is employed for multiclass classification, it serves as the activation function. The proposed system architecture is depicted in Fig. 1 and described in detail thereafter.

A. Datasets

To complete the implementation of the comparison of these transfer learning models a dataset of bird images available on Kaggle [31] is used for both the training and testing purposes. There are a total of 58388 images belonging to 400 species of birds for training purposes and 2000 images each of testing and validation images belonging to 400 species of birds.
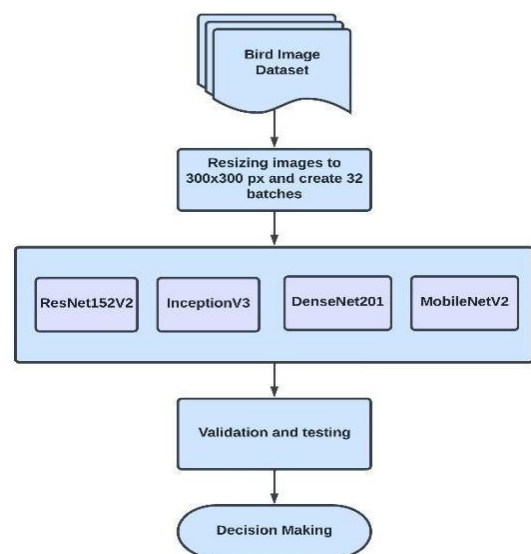

Fig. 1. System architecture.

B. Data Pre-processing

This mainly deals with preparing the data so that we can get maximum performance from the model after training the model using this data. All the images in the dataset must be in the same format for the model to be easier to train. For image pre-processing, we converted the images to 300x300px for standardizing the image dimensions. Then we created batches for training the model in batches. A batch size of 32 images was created. Batch size means the number of samples processed before the model is updated. This is needed because after training from batches expected results can be matched with actual results to calculate the error and this error is then used to update the algorithm to improve the model. All the images used are RGB images. So, the dimension of images used is (300,300,3).

C. Transfer Learning Models

Based on the literature survey, four transfer learning models are selected. These are ResNet152V2, InceptionV3, DenseNet201 and MobileNetV2. The ResNet152V2 and InceptionV3 models are selected because these are widely used and perform well. DenseNet201 is densely connected and thus takes more time to train but generally provides higher accuracy; and the MobileNetV2 is lightweight and has faster processing output.

1) DenseNet201: This is the feature extraction model. It is a convolutional neural network that is 201 layers deep. We can load millions of images to this pre-trained network which can classify the images into multiple object categories. Strong gradient flow, computational efficiency, and multiple different features are the advantages of this model.

2) ResNet152V2: This model is used for feature extraction. It is a pretrained model, so it gives higher accuracy in less time than traditional CNN. As the number of layers increases the training and testing rate also increases. To solve the problem of vanishing gradient residual block concept is introduced. The skip connection technique is used in this network. In this technique, it skips some layers in between. The advantage of this skip connection is if any layer in between changes the performance of architecture then this layer is skipped so performance increases. The ResNet model consists of a reshape layer, flattened layer, a dense layer having 128 neurons, a dropout layer, and the dense layer with SoftMax function which is used to classify the image.

3) InceptionV3: It is the deep learning model used for image classification which is based on CNN. It gives greater than 78.1% accuracy on the image datasets. It uses the transfer learning approach, so it gives good performance in classification. The model is made up of symmetric and asymmetric blocks which include convolutions, average pooling, max pooling, dropouts, concatenations, and fully connected layers.

4) MobileNetV2: It is a lightweight model for feature extraction which gives good performance on mobile devices. It is the convolutional neural network which is 53 layers deep. The model is based on residual structure. The model contains initial convolution layers with 32 filters followed by 19 residual bottleneck layers. The model can work in more than millions of images and the pre-trained network can classify images into about 1000 object categories.

D. Training

Each model is trained separately in a different notebook on Google Colab. An input layer and an output layer were added to the model for giving input and output along with a global average 2D pooling layer before the output layer. The Global Average 2D pooling layer takes a block of a tensor as input and calculates an average value of all the values across the tensor for each of the input channels. Pooling is necessary to reduce the dimensions, so it makes it easy for the model to train. The input layer accepts images of dimensions (300,300,3) and the output layer has 400 nodes for 400 species. The network is densely connected.

All the layers are frozen in the transfer learning model. The layers are frozen to reduce the training time and the activation function used is the same for all the models i.e., SoftMax as it is a multiclass classification. The learning rate is kept at 0.01. Based on the training time required to train the models and the size of the dataset used, 15 epochs were finalized to train the models. All these parameters are kept the same. This ensures that all the transfer learning models are fairly evaluated under similar conditions to give unbiased results.

E. Classification

A testing dataset of 2000 images is provided for testing the trained model. Accuracy is obtained to evaluate the performance of the models using the testing dataset. For prediction, the image which is to be classified is obtained from the user and it is first processed. It is converted into an image of 300x300px that is best suited for the model and then classification is done.

IV. RESULTS

The results are checked based on the accuracy provided by each model. The accuracy and loss curves are also plotted to better understand the training process of the model. This gives an idea about the accuracy and loss of the model over each epoch. The maximum observed accuracy is of ResNet152V2 model i.e., 95.45%. The plots of each model are given below.

As the models were built on the same parameters, the results obtained from the comparison can be considered to fairly evaluate the performance of these transfer learning models. The MobileNetV2 took less training time, and the model size is also small. The DenseNet201 has the best performance out of all and the InceptionV3 and ResNet152V2 performed well. ResNet152V2 displayed maximum accuracy but was having a large loss. These models can be compared based on various factors. The results obtained because of those factors are given below.

A. Performance Metrics

Given below are the performance metrics based on which the models are compared.

1) Accuracy: Accuracy in model training refers to the ability of the model to correctly classify input data. In other

words, accuracy is a measure of how often the model predicts the correct label for a given input. It is calculated as the ratio of correctly predicted samples to the total number of samples in the dataset. A higher accuracy score indicates that the model is better at predicting the correct output and is more reliable. However, it is important to note that accuracy is not the only measure of a model's performance.

2) Loss: Loss in model training refers to the difference between the predicted output of the model and the actual output. In other words, loss is a measure of how far off the model's predictions are from the true values. The goal of training a model is to minimize the loss function, which is typically a mathematical function that measures the difference between the predicted output and the actual output. This is done by adjusting the weights and biases of the network during the training process using techniques such as backpropagation. Lower loss values indicate that the model is better at predicting the correct output and is more accurate. However, it is important to find a balance between low loss values and overfitting, where the model becomes too specialized to the training data and performs poorly on new, unseen data.

3) Training time: Training time in model training refers to the amount of time it takes for the model to learn from the training data and adjust its weights and biases to minimize the loss function. The training time can depend on a variety of factors, such as the complexity of the model, the size of the training dataset, the available computational resources, and the hyperparameters used for training. The training process involves iterating through the training dataset multiple times (epochs) and adjusting the weights and biases of the network based on the feedback from the loss function. The goal is to find the optimal values for the weights and biases that minimize the loss and improve the accuracy of the model. In general, larger, and more complex models with larger datasets may require longer training times, while smaller and simpler models may converge faster. Efficient use of parallel processing resources can also help to reduce the training time.

B. Discussion

Given below are the accuracy graphs of the models.
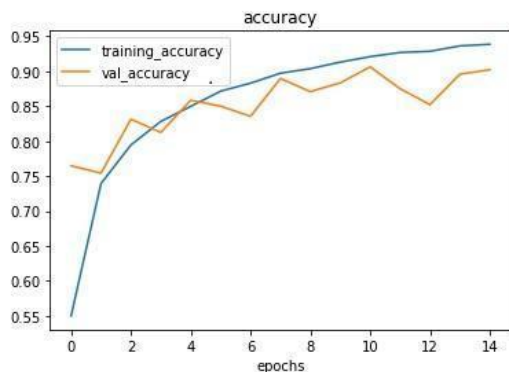

Fig. 2.   InceptionV3 model training and validation accuracy.
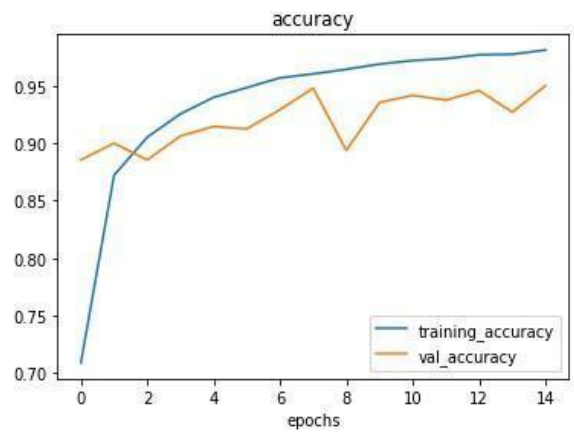

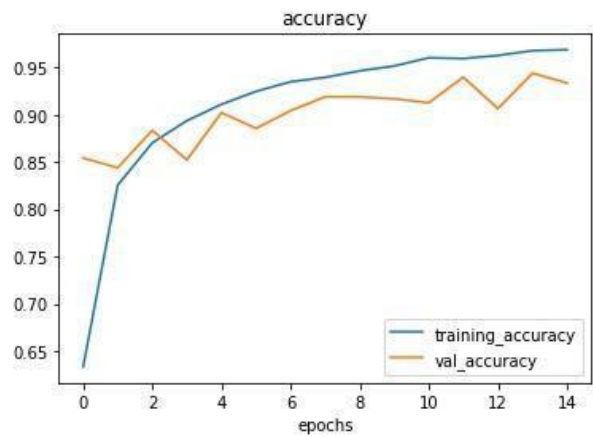Fig. 3.   ResNet152V2 model training and validation accuracy.

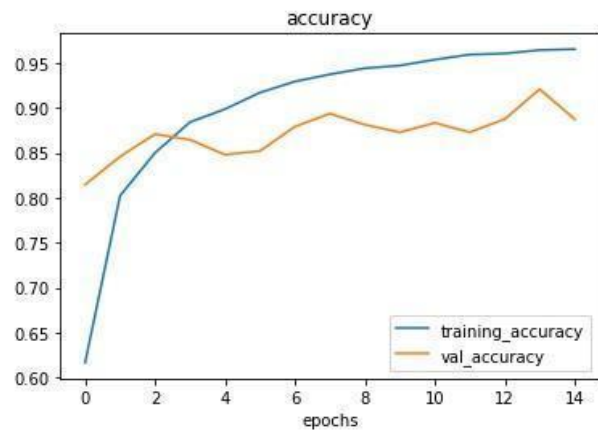Fig. 4.   DenseNet201 model training and validation accuracy.

Fig. 5.   MobileNetV2 model training and validation accuracy.

The steadiest training accuracy was observed in DenseNet201 from Fig. 4, and the other models performed almost similar in the training accuracy. The most stable validation accuracy was observed in DenseNet201 and the most unstable was observed in ResNet152V2 from Fig. 3 and InceptionV3 displayed in Fig. 2. From Fig. 5, it can be observed that the average training time of MobileNetV2 was 128 sec.

Given below are the loss graphs of the models.
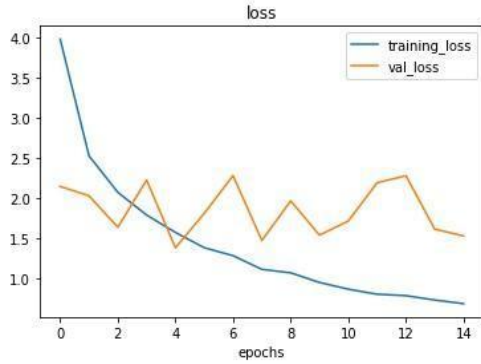

Fig. 6.   InceptionV3 model training and validation loss.
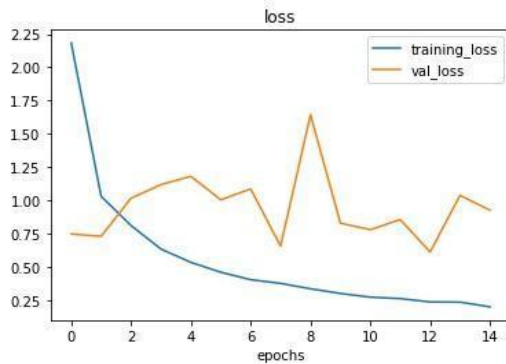

Fig. 7.   ResNet152V2 model training and validation loss.
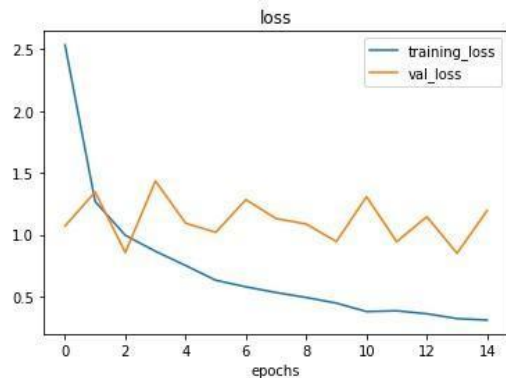

Fig. 8.   DenseNet201 model training and validation loss.
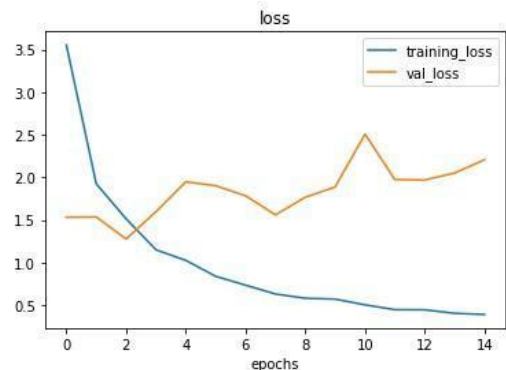

Fig. 9.   MobileNetV2 model training and validation loss.

The models performed similarly on training data by reducing the training loss over epochs, but the loss fluctuated during the validation. The most fluctuation was observed in InceptionV3 validation loss from Fig. 6 and the most stable among the models was observed in DenseNet201 from Fig. 8. Fig. 9 depicted that the validation loss in MobileNetV2 increased over epochs instead of decreasing. The training and validation loss of Resnet152V2 can be observed from Fig. 7.

Fig. 10 displays the comparative graph of average training time required per epoch for training the models.
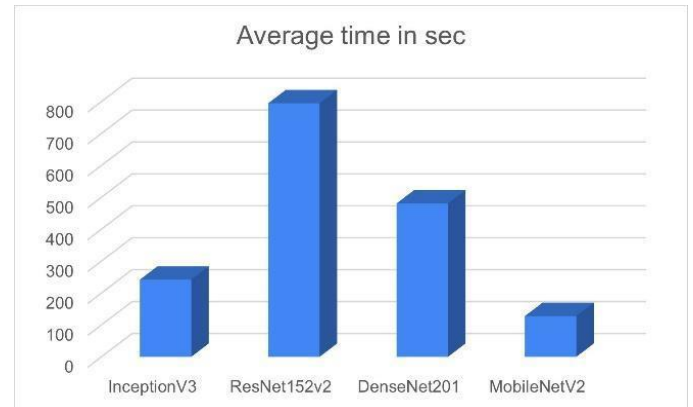

Fig. 10. Average training time per epoch comparison.

Fig. 10 compares the training time by taking the average of the time required by the model for each epoch. The ResNet152V2 required the highest training time with an average of 795 sec. It is more than the training time required by InceptionV3 and DenseNetV2 with average time 242 sec and 481 sec respectively. The MobileNetV2 is lighter and thus requires less time to train than the other transfer learning models compared in this study. The average training time of MobileNetV2 was 128 sec.

The overall result can be understood from the testing accuracy and the testing loss of the models as depicted in Table II.

Fig. 11 shows that ResNet152V2 provided the maximum accuracy with a loss of 0.883. But the DenseNet201 faced less loss than ResNet152V2 and provided accuracy like the ResNet152V2 model. Thus, DenseNet201 performed better than the other models on the bird image dataset with an accuracy of 95.09%.

TABLE II.   COMPARISON OF PERFORMANCES OF TRANSFER LEARNING MODELS

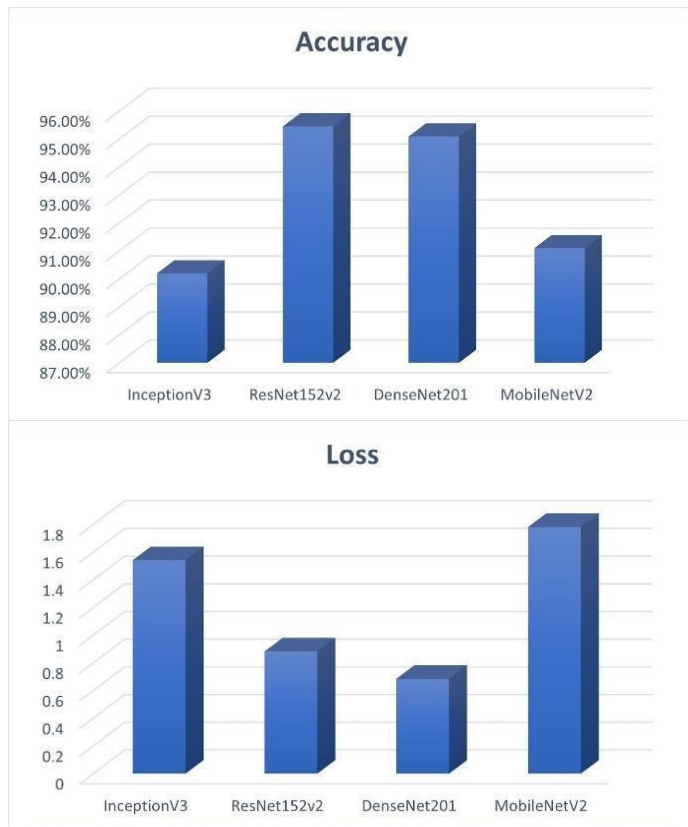| Model Name	Accuracy	Loss
InceptionV3	90.2%	1.542
ResNet152v2	95.45%	0.883
DenseNet201	95.09%	0.685
MobileNetV2	91.1%	1.782

Fig. 11. Accuracy and loss curves for comparing the models on testing data.

V. CONCLUSION AND FUTURE WORK

In this paper, various deep learning/ transfer learning models for bird species identification have been studied and compared. It has been understood that mostly the existing systems have used deep learning models or transfer learning models with smaller datasets. Therefore, an elaborate dataset consisting of 58388 [31] bird images of four hundred species has been identified for this work.

Four different transfer learning models namely InceptionV3, ResNet152V2, DenseNet201, and MobileNetV2 are implemented on the identified dataset. All these models were trained under similar conditions to get the best possible comparison between them. ResNet152V2 provided an accuracy of 95.45% which is more than the other 3 models. But it also faces more losses than DenseNet201. MobileNetV2 has the lowest training time, but the model accuracy is not as good as the other models. In conclusion, the best model among these models is DenseNet201. Even though its accuracy is a little less than the ResNet152V2 model, its loss is far less than other models. Thus, DenseNet201 is better than Resnet152V2.

Though the accuracy given by the implemented models are good, the models can be fine-tuned for better accuracy and better identification of birds. For fine-tuning, the number of epochs may be increased or some of the layers of the transfer learning models can be unfrozen. The same can be integrated with a mobile app, where common people would be able to upload captured images of birds and the app would provide details of the identified birds.

REFERENCES

[1] https://archive.epa.gov/water/archive/web/html/birds-initiatives.html.

[2] https://www.gov.uk/government/publications/wild-birds-licence-to-kill-or-take-for-conservation-purposes-gl40/list-of-endangered-woodland-birds.

[3] Mirugwe A, Nyirenda J, Dufourq E. Automating Bird Detection Based on Webcam Captured Images using Deep Learning. InProceedings of 43rd Conference of the South African Insti 2022 Jul 18 (Vol. 85, pp. 62-76).

[4] Hüseyin Gökhan Akçay, Bekir Kabasakal, Duygugül Aksu, Nusret Demir, Melih Oz, and Ali ¨ Erdoğan. Automated bird counting with deep learning for regional bird distribution mapping. Animals, 10(7):1207, 2020.

[5] Lynda Ben Boudaoud, Fr´ed´eric Maussang, Ren´e Garello, and Alexis Chevallier. Marine bird detection based on deep learning using high-resolution aerial images. In OCEANS 2019-Marseille, pages 1–7. IEEE, 2019.

[6] Biswas AA, Rahman MM, Rajbongshi A, Majumder A. Recognition of local birds using different cnn architectures with transfer learning. In2021 International Conference on Computer Communication and Informatics (ICCCI) 2021 Jan 27 (pp. 1-6). IEEE.

[7] M. A. Al-antari, P. Rivera, M. A. Al-masni, E. Valarezo, G. Gi, T. Y. Kim, H. M. Park, and T. S. Kim, "An automatic recognition of multiclass skin lesions via Deep Learning Convolutional Neural Networks," In Conference: ISIC2018: Skin Image Analysis Workshop and Challenge. 2018.

[8] L. Hu, and Q. Ge, "Automatic facial expression recognition based on MobileNetV2 in Real-time," In Journal of Physics: Conference Series, vol. 1549, no. 2, 2020.

[9] Hussain M, Bird JJ, Faria DR. A study on cnn transfer learning for image classification. InUK Workshop on computational Intelligence 2018 Sep 5 (pp. 191-202). Springer, Cham.

[10] Gao, Y., Mosalam, K.: Deep transfer learning for image-based structural damage recognition. Comput. Aided Civ. Infrastruct. Eng. (2018).

[11] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: IEEE CVPR 2016: Computer Vision and Pattern Recognition (2016).

[12] Raj S, Garyali S, Kumar S, Shidnal S. Image based bird species identification using convolutional neural network. International Journal of Engineering Research & Technology (IJERT). 2020;9(6):346.

[13] Nguyen, H., Maclaoup, S. J., Nguyen, T. D., Nguyen, T., Flemons, P., Andrews, K., ... & Phung, D. (2017, October). Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In 2017 IEEE international conference on data science and advanced Analytics (DSAA) (pp. 40-49). IEEE.

[14] Gavali, P., Mhetre, M. P. A., Patil, M. N. C., Bamane, M. N. S., & Buva, M. H. D. (2019) Bird Species Identification using Deep Learning.

[15] Rahman MM, Biswas AA, Rajbongshi A, Majumder A. Recognition of local birds of Bangladesh using MobileNet and Inception-v3. International Journal of Advanced Computer Science and Applications. 2020;11(8).

[16] S. Lee, M. Lee, H. Jeon, and A. Smith, "Bird Detection in Agriculture Environment using Image Processing and Neural Network," In 2019 6th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 1658-1663, 2019.

[17] J. Bankar, and N. R. Gavai, "Convolutional Neural Network-based Inception v3 Model for Animal Classification," International Journal of Advanced Research in Computer and Communication Engineering, vol. 7, no. 5, pp. 142-146, 2018.

[18] Wang Z, Wang J, Lin C, Han Y, Wang Z, Ji L. Identifying habitat elements from bird images using deep convolutional neural networks. Animals. 2021 Apr 27;11(5):1263.

[19] Guo, Q.H.; Jin, S.C.; Li, M.; Yang, Q.L.; Xu, K.X.; Ju, Y.Z.; Zhang, J.; Xuan, J.; Liu, J.; Su, Y.J.; et al. Application of deep learning in ecological resource research: Theories, methods, and challenges. *Sci. China Earth Sci.* 2020, *63*, 1457–1474.

[20] Islam S, Khan SI, Abedin MM, Habibullah KM, Das AK. Bird species classification from an image using VGG-16 network. InProceedings of

the 2019 7th international conference on computer and communications management 2019 Jul 27 (pp. 38-42).[Accessed: 04-Jun-2019].

[21] Huang YP, Basanta H. Bird image retrieval and recognition using a deep learning platform. IEEE Access. 2019 May 22;7:66980-9.

[22] Branson S, Van Horn G, Belongie S, Perona P. Bird species categorization using pose normalized deep convolutional nets. arXiv preprint arXiv:1406.2952. 2014 Jun 11.

[23] Ragib KM, Shithi RT, Haq SA, Hasan M, Sakib KM, Farah T. Pakhichini: Automatic bird species identification using deep learning. In2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4) 2020 Jul 27 (pp. 1-6). IEEE.

[24] Rezaeijo SM, Ghorvei M, Mofid B. Predicting breast cancer response to neoadjuvant chemotherapy using ensemble deep transfer learning based on CT images. Journal of X-ray Science and Technology. 2021 Jan 1(Preprint):1-6.

[25] Rosenberg KV, Dokter AM, Blancher PJ, Sauer JR, Smith AC, Smith PA, Stanton JC, Panjabi A, Helft L, Parr M, Marra PP. Decline of the North American avifauna. Science. 2019 Oct 4;366(6461):120-4.

[26] Shahabuddin G, Goswami R, Krishnadas M, Menon T. Decline in forest bird species and guilds due to land use change in the Western Himalaya. Global Ecology and Conservation. 2021 Jan 1;25:e01447.

[27] Díaz, S., Settele, J., Brondízio, E., Ngo, H., Gueze, M., Agard, J., Arneth, A., Balvanera, P., Brauman, K., Butchart, S., 2020. Summary for Policymakers of the Global Assessment Report on Biodiversity and Ecosystem Services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES).

[28] Burns F, Eaton MA, Burfield IJ, Klvaňová A, Šilarová E, Staneva A, Gregory RD. Abundance decline in the avifauna of the European Union reveals cross-continental similarities in biodiversity change. Ecology and Evolution. 2021 Dec;11(23):16647-60.

[29] Youth H. Birds in Decline. InVital Signs 2003–2004 2021 Apr 14 (pp. 82-83). Routledge.

[30] Sylvain Christin, Eric Hervet, and Nicolas Lecomte. Applications for deep learning in ecology. ´ Methods in Ecology and Evolution, 10(10):1632–1644, 2019.

[31] https://www.kaggle.com/datasets/gpiosenka/100-bird-species.