Bird Species Classification using Machine Learning

Final Project Report

1. Abstract

Bird species recognition is crucial for monitoring and conserving biodiversity, especially for endangered indigenous species. However, accurately identifying and tracking bird populations can be challenging due to visual variations and the complexity of species traits. This study employs supervised [4] and semi-supervised learning techniques to identify endangered bird species using decision trees and deep neural networks in machine learning. The research addresses the difficulties associated with classifying bird species by utilizing deep learning models, such as Convolutional Neural Networks (CNNs), to learn detailed characteristics and patterns from bird photos. The study highlights differences between machine learning techniques in bird species classification, including automatic feature learning, scalability, and transfer learning. The performance of the models is evaluated using metrics like accuracy, precision, confusion matrix, and t-SNE scatter plot. The findings demonstrate that CNN gives a better accuracy for image classification in comparison with Decision trees.

2. Introduction

2.1. Problem Statement

Birds play a critical role in global biodiversity, but many species are endangered. It is essential to accurately track and estimate bird populations, especially for local indigenous species. Unintentionally, wind farms have killed rare bird species in a number of nations. Due to differences in lighting and camera angles, it can be difficult to identify different bird species. Accurate models for species recognition are challenging to develop because of their variety in sizes, forms, colours, and physical traits. Through the use of supervised and semi-supervised learning techniques, this study attempts to identify bird species that are in danger of extinction by using decision trees and deep neural networks in machine learning [7, 10, 15].

The selection of the data set [3] is centred on the difficulty in identifying and classifying different bird species due to visual differences in size, form, colour, and other physical traits. The varying lighting conditions, various camera angles, and the complexity of the various bird species' visual traits are all difficulties in tackling this issue. Additionally, challenges are presented by the numerous bird species and the requirement for accurate identification. The selected method entails using deep learning models, such ResNet and CNNs, which are capable of learning detailed characteristics and patterns from the photos, to overcome these issues.

There are a number of benefits and drawbacks to the current machine learning systems for classifying bird species. Positively, automatic feature learning is made possible by machine learning techniques, which eliminates the necessity for manual feature engineering. By doing so, the models are able to recognise intricate patterns and hierarchies in photos of birds, potentially increasing classification accuracy. Additionally, when trained on substantial and varied datasets, machine learning models like Convolutional Neural Networks (CNNs) can attain great levels of accuracy. They are suited for use in real-world contexts since they are resilient to changes in lighting, backdrop, stance, and scale within the collection of bird species. Another benefit is transfer learning, which makes it possible to use pretrained models and increase performance with little training data. At last, scalable machine learning methods can handle huge datasets with a variety of bird species and photos. The need for a sizable amount of labelled training data, the computing resource-intensive nature of training, the lack of interpretability caused by complex models, the susceptibility to noisy data, and the danger of overfitting when training on small datasets are obstacles. However, machine learning offers a potential method for classifying bird species by carefully addressing these limitations. [18]

The use of pre-training on a dataset containing several bird species classes, followed by fine-tuning and training from scratch using stochastic gradient descent (SGD) optimisation, is one potential approach to overcoming the difficulties that we have used In our project. A variety of criteria, including accuracy, precision, and confusion matrix analysis, have been used to assess the models. Plots of the model's performance that demonstrate diminishing loss and rising accuracy over time throughout training has been generated to shed light on the models' capacity for learning. Secondly, we have utilized the Decision Tree algorithms for both supervised and semi-supervised learning models [5]. Decision Trees are powerful and interpretable machine learning algorithms that partition the data based on feature values, enabling effective classification. By employing Decision Trees, we aim to leverage their capabilities in capturing complex patterns and relationships within the bird species dataset.

To evaluate and compare the performance of these models on 2 different datasets, we employed two key evaluation techniques: the confusion matrix and t-SNE scatter plot. These evaluations have guided us to determine the most effective approach for achieving improved accuracy and robust classification results.

2.2. Related Work

In Taiwan, one of the literature suggested using the transfer-based learning method using Inception-ResNet-v2 [13] to detect and classify the bird species endemic to Taiwan and to distinguish them from the other object domains [10].

This paper explores the implementation of a Convolutional Neural Network (CNN) for bird species identification. The CNN architecture includes an input layer for grayscale images, an output layer for binary or multi-class labeling, and hidden layers consisting of convolutional layers, ReLU layers, pooling layers, and a fully connected neural network. These components work together to extract features and make accurate predictions about the bird species. [17]

This study focuses on identifying bird species using the SVM classifier and decision tree techniques. In order to categorise diverse bird species, the SVM classifier uses machine learning to identify the best hyperplane that divides them into distinct groups. The decision tree method, on the other hand, bases judgements on the characteristics of the bird species using a tree-like structure. This study's goal is to investigate how well these two classifiers perform in correctly identifying and categorising various bird species. [16]

The research utilizes the 285-birds dataset, which consists of 43,780 digital images with 285 class labels. Transfer-learning techniques are employed, using pre-trained weights from models like AlexNet and ResNet34 trained on ImageNet. The research shows that the ResNet model with sparse regularization achieves robust performance in recognizing a large number of wild birds, suggesting its suitability for large-scale bird recognition systems. [9]

After reviewing the existing literature, we have chosen to employ the ResNet [19] Model for our bird species recognition task. To optimize the model, we utilized the SDM optimizer and incorporate the ReLu activation function within the Convolutional Neural Network (CNN) architecture. Additionally, we have incorporated the Decision Tree algorithm into our semi-supervised model and supervised model.

3. Methodologies

3.1. Datasets

We have selected the dataset from the Kaggle, which consists of 25 Indian Bird species. The chosen dataset in this case is balanced. Each photo was hand-picked and taken from the eBird Platform. A total of 22.6k photos, representing all 25 distinct bird species, have been painstakingly labeled by the author. Here is the dataset's detailed specification.

The given data set consists of around 925 images of 25 different kinds of species of birds. Images are taken from the bird conservation community Platform and most of them consist of 1200*800 resolution. The cleaning process of the data set has been done with a Python script to generate the properly balanced dataset for training, validation, and testing. here is the detailed specification of the dataset.

Label	No. of Observations
Asian Green Bee-Eater	924
Brown-Headed Barbet	924
Cattle Egret	918
Common Kingfisher	924
Common Myna	926
Common Rosefinch	919
Common Tailorbird	918
Coppersmith Barbet	778
Forest Wagtail	924
Gray Wagtail	926
Hoopoe	924
House Crow	924
Indian Grey Hornbill	922
Indian Peacock	919
Indian Pitta	629
Indian Roller	926
Jungle Babbler	930
Northern Lapwing	930
Red-Wattled Lapwing	930
Ruddy Shelduck	925
Rufous Treepie	860
Sarus Crane	930
White Wagtail	930
White-Breasted Kingfisher	930
White-Breasted Waterhen	930

Table 1. Labeled Data Observations

We have used the Python script that preprocesses the dataset to eliminate the noisy data and resize the data in order to clean and prepare the dataset for our model training. To standardize the image inputs for the neural network, the original images have been resized to 224*224 while preserving the original aspect ratio. Some photos with incorrect sRBG formats were discovered during preprocessing and we have removed them from the dataset. We have also split the dataset into 3 subset: training (70%), testing (15%) and validation (15%) dataset. To address the issue of lengthy training times and to focus on a more manageable subset of bird species, we decided to reduce our dataset to 10 bird species classes. This reduction in the dataset size enabled us to mitigate the training time while still maintaining a reasonable level of accuracy. We have used 10 classes total of 8877 images for ResNet DL Model, for Decision Tree we have used 3 classes of bird species total consisting of 2767 images.

3.2. Decision Tree Model

For the Decision tree classification model, we have employed certain methods like feature extraction, hyperparameter tuning methods also loading and labeling the image dataset after preprocessing. In the pre-processing, the images are resized from 224*224 to 128*128 resolution. We have used certain feature extraction methods: The HOG features [11] to capture the shape and structure of an image by computing gradients in various directions, while the LBP features describe the texture and patterns in an image by encoding the local pixel variations. Gabor filters use a variety of orientations and frequencies to convolve the image in order to extract texture information. These features are taken out of the image once it has been converted to grayscale. In order to generalise the model, we have also taken into account the image's colour histogram. Due to the lengthy nature of this feature extraction technique, we have whittled down our dataset to just 3 classes, totaling 2767 observations.

For the hyperparameter tuning in both supervised and semi-supervised learning, we have used the 'GridSearchCV' from the sci-kit-learn using the hyperparameter grid. The Hyperparameter grid contains parameters like max_depth, max_features, min_samples_split, and min_samples_leaf. By just providing the list of this parameter, this method automates hyperparameter tuning by trying every possible combination and return the best parameter combination that we can use to train our decision tree model using the best_estimator_ attribute of the GridSearchCV object.

For semi-supervised learning, a pseudo-labeling approach is used. It starts by randomly selecting a portion of the training data as labeled and the remaining as unlabeled. The decision tree classifier is trained using the labeled data, then it iteratively follows these steps, predicts (pseudo-labels) the unlabeled data using the trained classifier. Filters high-confidence pseudo-labeled data based on specified confidence thresholds. Mixes the high-confidence pseudo-labeled data with the labeled data. Retrains the classifier using the updated labeled dataset. Updates the labeled and unlabeled datasets for the next iteration. This process is repeated for a specified number of iterations, generally 5-10 iterations.

3.3. DNN Model

We utilized ResNet [19], a highly influential deep learning model developed by Microsoft Research, in our bird classification project. ResNet effectively addresses the challenge of vanishing gradients in deep neural networks by incorporating residual connections that enable the direct flow of gradients from earlier layers to later layers. This architectural design allows for the training of deep networks without suffering from performance degradation.

Our ResNet model was trained from scratch using stochastic gradient descent (SGD) [8] optimization [6]. SGD is a popular optimization technique in deep learning that approximates the true gradient by computing gradients on randomly selected mini-batches of training data. This approach ensures efficient updates of the model's parameters, enabling effective learning. The architecture of the ResNet model consists of multiple residual blocks. Each residual block contains a set of convolutional layers, followed by batch normalization and ReLU [2] activation. The input of the block is added to the output through a skip connection, allowing the network to learn residual mappings rather than directly learning the underlying mapping. This architecture helps in training deeper networks by alleviating the degradation problem. The ReLU activation function is chosen for its simplicity and ability to introduce nonlinearity, which is crucial for capturing complex patterns in the data. [6]

In terms of computational complexities, the training and validation phases of the selected models involve two main aspects: wall clock time for one-epoch training and the number of floating-point operations (FLOPS) calculation. In terms of computational complexities, the wall clock time for one-epoch training is reported as 778.73 seconds for the first iteration, which may involve additional initialization and setup time. In the subsequent iterations, the average training time is reduced to 44 seconds. The exact training time can vary depending on the hardware used, such as GPU acceleration, and the complexity of the model and dataset. The number of FLOPS (floating-point operations) calculation gives an estimate of the computational complexity of the model. The reported value is approximately 3,666,993,152 FLOPS. The FLOPS calculation depends on the specific architecture, layer configurations, and input size of the model.

During the training process, our reduced dataset achieved an accuracy of 95.44%. For testing, the model yielded an accuracy of 86.36% with 10 epochs, indicating its capability to generalize well to unseen data, and for validation it has achieved an accuracy of 87.18%.

3.4. Optimization Algorithm

We have used stochastic gradient descent (SGD) for optimizing the training of CNNs, which involves breaking up the labelled dataset into smaller batches and updating the CNN's weights and biases in accordance with the gradients of the loss function. Hyperparameters used for CNN include, learning rate 0.001, batch size of 32, which is optimized by SGD. Figure 1, shows loss function with every epoch. Metrics considered during the comparison are accuracy, precision, recall and F1 score. We are using Rectified Linear function as the activation function and the neuron is called Rectified Linear Unit (ReLU).

ReLU's use aids in avoiding the exponential growth of compute needed to run the neural network. The "vanishing gradient" problem is avoided by ReLUs, with which even with high input values to the activation function, backpropagation of the error and learning can continue.

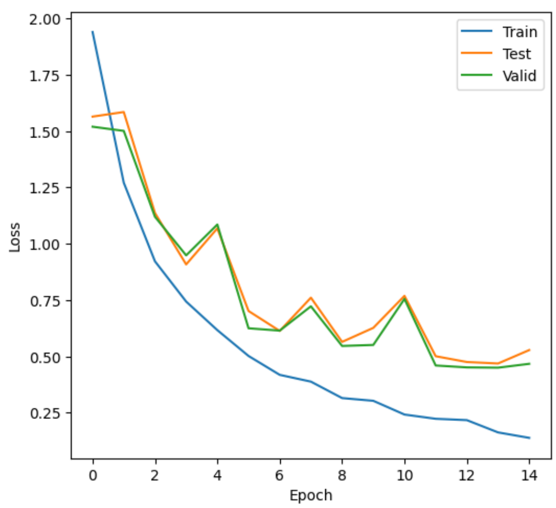

Figure 1. Loss functions with epochs.

GridSearchCV is the optimisation approach we have used for Decision trees with both supervised and semi-supervised. It methodically looks for the best possible set of hyperparameters within a given parameter grid. Accuracy and cross-validation are the metrics used by GridSearchCV to evaluate different combinations of hyperparameters.

The hyperparameters used for supervised decision tree are [10,20,30,40] as Max depth, [2,5,10] as minimum sample split, [1,2,4] as minimum sample leaf, and the one used for semi-supervised learning are [20,30,40] as Max depth, [2,5,10] as minimum sample split, [1,2,4] as minimum sample leaf, which are optimized by the GridSearchCV to maximise accuracy. The labelled data is split into training and validation subsets, with decision tree models being iteratively trained with various hyperparameter values and their effectiveness being assessed on the validation sets.

4. Results

4.1. Experiment Setup

Hardware Setup : Intensive processing unit is needed for the model's training. In order to employ cloud computing and a lot of processing power for this project, we used Google Colab [1]. The hardware setup provided by Google for this project is described in detail here.

Name	Specification
GPU	Tesla T4 - 15.109 GB
CPU	Intel(R) Xeon(R) CPU @ 2.30GHz
CPU Frequency	2000.178 MHz
RAM	12.7 GB
Disk Space	107.7 GB

Table 2. Google Collab Hardware Configuration

Hyper-parameter : The CNN (Convolutional Neural Networks) model's and Decision Tree's performance are greatly influenced by the hyper-parameters chosen for training. We need to take into account the type of dataset, its size, as well as its dimensions and noisy data while choosing the optimised hyperparameter. In order to choose the optimum hyperparameter, we must also take computational resources into account. Based on its architecture, each model has a separate set of hyperparameters.

Learning Rate	0.001
Batch Size	32
Epochs	15
Input size of image	224*224*3
Optimizer	SGD

Table 3. Hyper-Parameter for ResNet CNN Model

In decision trees, different hyperparameters are used to control the model's behavior. For example, max_depth limits the tree's depth, while min_samples_split and min_samples_leaf set the minimum number of samples required for splitting and leaf nodes. The max_features parameter determines the number or portion of features considered for splitting. These hyperparameters help prevent overfitting and improve performance.

The performance of Decision Tree (supervised and semi-supervised) models with the ResNet CNN model, has been evaluated using metrics such as accuracy, recall, precision,

Max Depth	[10, 20, 30, 40]
Min Sample Leaf	[1, 2, 4]
Min Sample Split	[2, 5, 10]
Cross Validation	2
Max Feature	[2, 3]

Table 4. Hyper-Parameter for ResNet CNN Model

and F1-score. Accuracy measures overall prediction correctness, recall assesses the identification of positive instances, precision evaluates the correct positive predictions, and the F1-score balances precision and recall. Moreover, Graph plotting techniques such as the Confusion Matrix and t-SNE were used to visually assess the performance of a model. The Confusion Matrix provides a summary of classification accuracy and error patterns, while t-SNE helps visualize data clusters and class separability in a lower-dimensional space. These visualizations aid in understanding the model's performance more intuitively.

4.2. Main Results and Ablative Study

The performance of the ResNet model in CNN was evaluated using metrics such as Loss, Recall, Precision, and F-Score. The model showed a decreasing loss, increasing accuracy, and improved precision over each epoch. The F-Score values indicated a good balance between precision and recall. Overall, the model exhibited continuous improvement in its ability to accurately classify bird species.

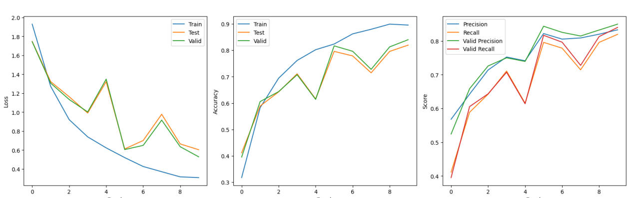

Figure 2. ResNet Model Performance over each epoch with respect to Loss, Recall, Precision and F-Score

The performance of the trained models, particularly in decision tree classification, was evaluated using two datasets: one with 10 clThe performance of the ResNet model in CNN was evaluated using metrics such as Loss, Recall, Precision, and F-Score. The model showed a decreasing loss, increasing accuracy, and improved precision over each epoch. The F-Score values indicated a good balance between precision and recall. Overall, the model exhibited continuous improvement in its ability to accurately classify bird species.asses and another with 3 classes. The performance metrics were compared between the testing and validation sets. For the dataset with 10 classes, it was observed that the performance metrics were higher in the testing set compared to the validation set. This suggests that the decision tree model achieved better generalization and accuracy when predicting the unseen testing samples. However, the scenario was different for the dataset with 3 classes. Here, the performance metrics showed that the validation set outperformed the testing set. This could indicate that the decision tree model may have overfit to the training data, resulting in lower performance on the testing set.

Model	Dataset Size		Accuracy	Precision	Recall	F1 - Score
Resnet CNN Model	10 Class 8877 Images	Test	86.36	87	86	86.49
		Valid	87.16	88	87	87.49
Decision Tree Supervised	10 Class 8877 Images	Test	26.75	26.50	26.26	26.25
		Valid	25.79	25.23	25.14	25.10
Decision Tree Supervised	3 Class 2767 Images	Test	48.92	50.11	48.92	49.05
		Valid	51.08	51.04	51.08	50.80
Decision Tree Semi - Supervised	3 Class 2767 Images	Test	45.34	45.39	45.34	45.36
		Valid	54.44	54.51	54.47	54.48

Figure 3. Comparison of Performance Matrices in 3 models

As the Ftigure 3 demonstrates, With the image dataset, the decision tree does not perform well. By reducing the number of classes, we can improve the performance of the matrix. The main reason for this is that training the model for 10 classes increases the likelihood of overfitting, whereas training on 3 classes only reduced the overfitting, leading to an improvement in performance. Performance can be improved by choosing the optimum hyperparameter. The decision tree's depth was extended to 100 during the initial phases, when we classified 25 different bird species using the original dataset. As a result, the model's performance was 22%.

In the ResNet CNN model, the confusion matrix [14] is used to analyze the performance of the trained model. The heat map illustrates the predicted labels in relation to the true labels. From the graph, we observe that our model is well-fitted to the training data, indicating its ability to accurately classify unseen and unlabeled data. This is further supported by the high accuracy shown in the confusion matrix, indicating that our model performs well in identifying different classes of data. Overall, these results demonstrate the effectiveness and reliability of our model in classifying and predicting labels for unseen and unlabeled instances.

In the supervised model of decision tree, the accuracy on the validation dataset is notably higher than on the testing dataset. These results demonstrate the model's strong performance in both training and generalization tasks by analayzing given Figure 5.

As per the given Figure 6, The performance of the model on the validation dataset is superior to that on the testing dataset. However, even though the performance on the testing dataset may be comparatively lower, it still outperforms the supervised learning model. This suggests that the model

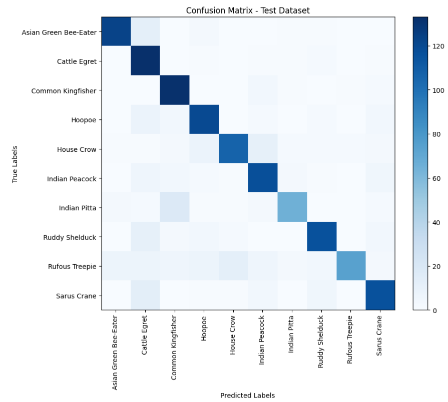

Figure 4. Confusion Matrix to visualize the difference of predicted and true label for ResNet

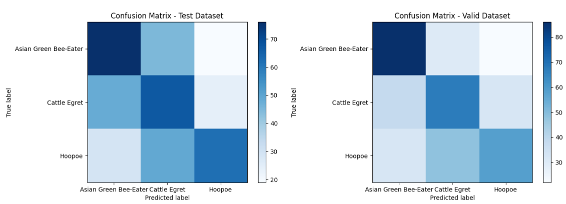

Figure 5. Confusion Matrices to visualize the difference of predicted and true label for supervised model on training and validation dataset

has a higher level of generalization and is more effective in handling unseen data.

Since the decision tree's training and hyperparameter tuning phases took longer when the batch size was 32 and the image resolution was 224*224, we increased the batch_size to 128 and decreased the decision tree's image resolution to 128*128 in order to extract features from the image. This significantly reduced training time, but the dataset's slight overfitting was still visible. We attempted to train the semi-supervised model on the 10 bird species using 8877 observations, however it took too long to train and put a strain on the Google Collab's computer resources.

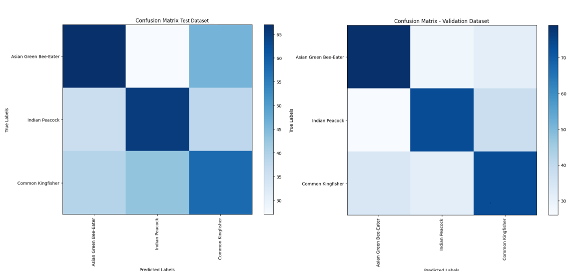

Figure 6. Confusion Matrices to visualize the difference of predicted and true label for semi-supervised model on training and validation dataset

t-SNE [12] is a technique used to visualize high-dimensional data in a lower-dimensional space. It helps us observe how data points with the same class label cluster together and enables the identification of any outliers within the dataset. Upon comparing the provided t-SNE plots, it becomes evident that the CNN model outperforms both the supervised and semi-supervised models. The CNN model demonstrates superior performance in terms of data clustering and the identification of patterns, showcasing its effectiveness in capturing complex relationships within the dataset.

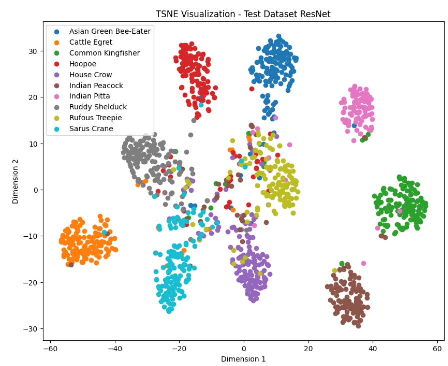

Figure 7. ResNet Model t-SNE

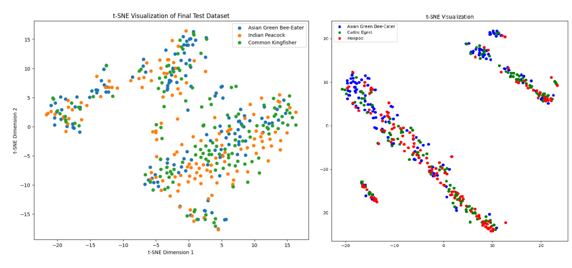

Figure 8. Comparision of t-SNE Plot on Supervised and Semi-supervised model

Initially, we trained on a dataset with 25 classes and 22.6k observations. To expedite training, we reduced it to 10 bird species. No hyperparameter tuning was done. By implementing data augmentation, accuracy improved from 11% to 95%. Data augmentation involves creating modified copies of existing data to expand the training set.

In conclusion, CNNs are superior to decision trees for image classification because they can automatically recognise intricate patterns and directly extract spatial connections from images. While, decision trees depend on manually created features and might not attain the same level of accuracy, especially for huge datasets. Therefore, CNNs are favoured over decision trees for tasks like bird species identification, where correct classification is essential.

References

[1] Google colaboratory. Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. https://colab.google/. 4

[2] baeldung. How relu and dropout layers work in cnns. *Baeldung CS*, page 1, 2023. 3

[3] Arjun Basandrai. 25 indian bird species with 22.6k images, 2023. Dataset of 25 bird species for image classification. 1

[4] Ayoub Bensakhria. Bird classification system based on osteological measurements using supervised machine learning algorithms (a methodology proposal), 12 2022. 1

[5] Baowen Qiao; Zuofeng Zhou; Hongtao Yang; Jianzhong Cao. Bird species recognition based on svm classifier and decision tree. *2017 First International Conference on Electronics Instrumentation & Information Systems (EIIS)*, 2018. 1

[6] Anshu Raina Chirayu Garg, Abhinav Garg. Analysis of gradient descent optimization algorithms on resnet, 2017. Department of Computer Sciences, University of Wisconsin-Madison. 3

[7] Piyush Bidwai; Vaibhav Mahalle; Eshan Gandhi; Sharda Dhavale. Bird classification using deep learning. *International Research Journal of Engineering and Technology (IRJET)*, 7, 2020. 1

[8] Ayush Gupta. A comprehensive guide on optimizers in deep learning, 2021. Blog on Deep Learning Optimizers - Stochastic Gradient Descent (SGD), Adam. 3

[9] M. Muazin Hilal Hasibuan, Novanto Yudistira, and Randy Cahya Wihandika. Large scale bird species classification using convolutional neural network with sparse regularization. In *Proceedings of the 2022 Brawijaya International Conference (BIC 2022)*, pages 651–663. Atlantis Press, 2023. 2

[10] Yo-Ping Huang and Basanta Haobijam. Recognition of endemic bird species using deep learning models. *IEEE Access*, PP:1–1, 07 2021. 1, 2

[11] Susanto Kumar Ghosh; Md. Rafiqul Islam. Bird species detection and classification based on hog feature using convolutional neural network. *Communications in Computer and Information Science (CCIS)*, 1035, 2019. 3

[12] Geoffrey Hinton Laurens van der Maaten. Visualizing data using t-sne. *Journal of Machine Learning Research 9 (2008) 2579-2605*, 2008. 6

[13] Md. Mahbubur Rahman; Al Amin Biswas; Aditya Rajbongshi; Anup Majumder. Recognition of local birds of bangladesh using mobilenet and inception-v3. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 11(8), 2020. 2

[14] Sarang Narkhaede. Understanding confusion matrix. 2018. 5

[15] Benedict Neo. Bird species classification with machine learning. 2022. 1

[16] Baowen Qiao, Zuofeng Zhou, Hongtao Yang, and Jianzhong Cao. Bird species recognition based on svm classifier and decision tree. pages 1–4, 06 2017. 2

[17] Dharaniya R, Preetha M, and Yashmi S. Bird species identification using convolutional neural network. 11 2022. 2

[18] Sathi Chatir Vinay Reddy Sathi Veerraghava Reddy. Bird species identification from image, 2022. Project Report submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering's Degree in Computer Science and Engineering. 1

[19] Yucheng Li; Hua Zhou; Yili Zhao. The fine-grained recognition of bird images based on joint semantic components and resnet. *2022 5th Internation Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2022. 2, 3