2024 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT) | 979-8-3503-8592-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONECCT62155.2024.10677184

Multiclass Bird Species Identification using Deep Learning Techniques

Pruthvi Darshan S S[1], Jayanth M L[2] and Sangeetha V[3]
Department of Computer Science & Engineering
Ramaiah Institute of Technology
Bengaluru, Karnataka, India
Email: [1]darshupruthvi007@gmail.com, [2]mljayanth07@gmail.com, [3]drsangeethav@msrit.edu

***Abstract*—Identification of bird species is a necessity for conserving a large ecosystem that is rapidly declining due to climate change and human destruction. Earlier identification depended on the manual efforts of human resources, who were researchers and ornithologists, which was time-consuming, and there was a chance of error and misclassification. In order to address this issue, this paper utilizes deep learning technology to identify 525 species of birds using ResNet50, MobileNetV3, EfficientNet-B0, and Wide-ResNet50V2. This analysis focuses on assessing the model's accuracy, training, and validation loss/accuracy and measuring the performance of these models through evaluation metrics and testing accuracy to determine their effectiveness in species differentiation tasks. Results demonstrate that EfficientNet-B0 outperforms other models, achieving an accuracy rate of 99.09% trained over a dataset of 84635 images. This approach holds promising results for accurate species identification for researchers and ornithologists in tracking and preserving rare bird species for ecosystem conservation efforts and advancing the field of computer vision & deep learning in biodiversity monitoring and species identification.**

***Index Terms*—Bird Species, CNN, EfficientNet-B0, MobileNetv3, PyTorch, ResNet50, Wide-ResNet50V2**

I. Introduction

Birds are feathered superstars in the ecosystem, constantly reminding the natural world around people. They are seen travelling overseas featured in myths, legends and folklore across cultures. Millions of humans worldwide enjoy most of their time birdwatching. It provides a calming and enjoyable daily activity allowing people to observe birds in their natural habitat. They offer companionship and entertainment. They play an important role in agriculture by controlling insect populations and pollinating plants. Ravens, vultures and other scavengers clean up the environment by consuming carrion (dead animals), returning vital nutrients back into the ecosystem and preventing the spread of diseases.

The classification of bird species is a fundamental task in conservation, ecology, and ornithology. With over an estimated 15000 to 18,000 species worldwide [1], identification relies on manual observation or field guides, which are limited in their scalability and accuracy and also very time-consuming, error-prone, and labor-intensive. Not just images but also sounds and videos can also be used in bird identification. This paper leverages Convolutional Neural Networks (CNNs) [2], which is one of the deep learning techniques, that emerged as powerful tools to address this challenge with fewer computational difficulties. Bird populations can be served as indicators of ecosystem health, monitoring changes in their distributions and abundances can provide valuable insights into environmental changes. It can help in planning conservation strategies to prioritize habitats for resource allocation and protection. Understanding bird species interactions, their migration patterns, and habitat preferences, which rely on accurate species identification. In recent years, there have been a surge in the availability of publicly accessible datasets of different bird species, necessitating scalable and efficient classification algorithms.

Here, this research involves analyzes on bird image datasets of 525 species consisting over 90,000 images which is a very high image quality dataset where each bird occupies at most 50% of pixels of image. All these are original images captured from either a digital camera or smartphones with a male and female bird ratio of 80% and 20% respectively and not created by augmentation. From these images, research utilizes feature engineering and image enhancement techniques to effectively identify birds. By leveraging the use of PyTorch framework to train powerful classification models namely, ResNet50, MobileNetv3, EfficientNetB0 and Wide ResNet50v2 with custom training methods, and finally evaluate its credibility by assessing with few evaluation metrics to measure its performance.

This paper is organized into five sections. Section 2 discusses related work, specifically on birds species identification and deep learning techniques proposed to face this approach. Section 3 focuses on the proposed methodology for classification of birds species using deep learning models. Results and discussion are reported in section 4 and conclusions are drawn in section 5.

II. Literature Review

Many bird species are rapidly getting extinct due to either climate changes or habitat destruction caused by human activities. For an effective conversation planning, it is important to accurately know the species distributions and factors contributing to a specific region's biodiversity. In earlier periods, people used to listen to the audio of birds to recognize different species, due to external factors like wind, background noises, other birds and insects sound made it difficult and unreliable to effectively classify. Researchers and scholars found image classification more reliable and with the rise of deep learning methods, extracting features from birds and classification made a major breakout in this field of research. Later on ornithologists and people were getting benefits using these systems. Video processing was also a reliable method, but due to noise, difficulties in appearance between species due to lighting conditions, differences in shapes and resulting in poor accuracy. People find images better than audio or video processing techniques. With image classification techniques, they could possess deeper insights on the migration and adaptable behavior of different bird species in ecosystem on different climate conditions and seasons.

Ian Ting et al. [3] experimented by training multiple models to classify 15 classes of bird species which can be rarely found at Malaysia and in other parts of the world, and studied a brief performance measure. The DenseNet169 model was very effective and outperformed ResNet152V2, InceptionV3, and DenseNet121 for their dataset comprising over 35000 images.

Hany, Ayman and Atia [4] studied pre-trained deep learning models in order to identify bird species utilizing CNN classification and bottleneck feature extraction methods. Their performance metric includes recall, precision and F1-score to evaluate and assess the model performance. Their results show high levels of accuracy among ResNetv2 and InceptionV3 models, and VGG19 had a comparably low accuracy which made it unreliable for their dataset.

Neeli et al. [5] proposed a method for identification of bird species using EfficientNet-B0 architecture and integrate it to an application which automatically classify bird species from image, and explored other aspects such as feature engineering and fine-tuning to optimize the model architecture performance after training with 84,635 images. Their research was indicated a clear breakdown of results on each of their metric evaluation and how the model predicted for random images on bird species.

Duta and Behdad [6] proposed a solution for visual recognition in noisy environments for birds classification. They incorporated data augmentation for bird images to increase its diversity and they trained model by utilizing deep learning neural networks using a noisy robust loss function which will penalize the model for false predictions caused by noise. Their method achieved superior performance when compared with legacy methods and good accuracy in both clean and noisy environments. However, they still had issues with few inter-species variance classification which implies room for further improvements.

Identification of bird species can be done manually by individuals who were familiar with birds, such as ornithologists and people who have limited knowledge of birds, cannot identify just by looking at the images of birds. Even experts are not so familiar with all of the bird species, and employ deep learning models to identify and classifying bird species accurately and compare of these models for better accuracy to reduce time and human effort. Existing studies have demonstrated the effectiveness of Convolutional Neural Networks (CNNs), LSTM, and YOLO architecture in accurately classifying bird species on image data. However, a significant research gap exists in the availability of both limited and comprehensive datasets encompassing a diverse range of avian taxa. While some studies focus on specific regions or taxonomic groups, there is a lack of standardized datasets that capture the global diversity of bird species. This research gap is crucial for developing deep learning models, given the importance of robust training datasets for accurate and generalizable models.

III. Proposed Methodology

A. PyTorch Pretrained Models

PyTorch [7] is a flexible framework with extensive collection of pretrained models make it a popular choice for deep learning research and application development. These are pretrained neural network architectures models that have been trained on large scale datasets for specific tasks, such as object detection, classification and natural language processing. Using transfer learning technique, a model trained on one task can be adapted or transferred to another related task with limited labeled data. This approach can significantly reduce effort to label the data and the amount of labeled data required for training and improving model generalization. From collection of its popular architectures, this research employs ResNet50, MobileNetv3 small, EfficientNet-B0, and Wide ResNet50V2 to effectively classify bird species.

B. Research Setup

PyTorch provides flexible and dynamic approach while building and training neural networks, offering rich ecosystem of libraries and tools for various tasks, including data preprocessing and loading, model building, optimization and deployment making it popular in the field of deep learning. Tensors in PyTorch can seamlessly be moved across CPU and GPU for accelerated computation. Autograd module enables automatic differentiation of tensor operations, facilitating more controlled gradient-based optimization methods for training neural networks, and allowing for the implementation of custom loss functions and optimization algorithms.

Fig. 1 illustrates images from the dataset which were randomly plotted with 15 different bird species. With custom training and validation methods, the dataset is distributed as 84635 images for training, 2625 images for validation, 2626 images for testing, and transform each of these images to tensors of size 224x224. The models are trained for 32 batch size, 20 epochs, LambdaLR scheduler to specify custom learning rate function to adjust it during training, initial learning rate of 0.01 with Adam optimizer to maintain separate learning rates for each parameter, and calculates individual adaptive learning rate for different parameters, and calculating cross entropy loss between predicted class scores and target class labels. This setup is visualized with the system architecture of proposed model shown in the Fig. 2.

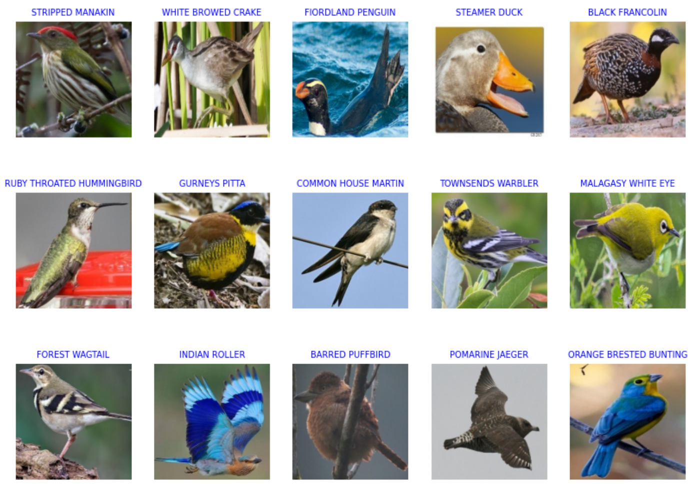

Fig. 1: Random sample of 15 different bird species from dataset

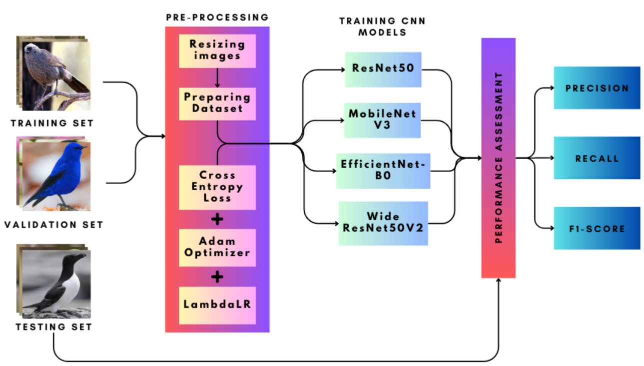

Fig. 2: System architecture of Proposed model

This research employs precision, recall and F1-score evaluation metrics to compare and assess these model to facilitate informed conclusions and discussions.

- Confusion matrix is used to assess the performance of classification model and allows to visualize it by showing the number of correct & incorrect classifications instances.
- True Positives (TP) are the instances that are predicted correctly as belonging positive class.
- False Positives (FP) are the instances that are incorrectly predicted as belonging to the positive class when they actually belong to a negative class.
- True Negatives (TN) are the instances that are predicted correctly as belonging to the negative class.
- False Negatives (FN) are the instances that are incorrectly predicted as belonging to the negative class when they actually belong to a positive class.
- Precision is the metric used to measure the proportion of true positives out of all positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

- Recall is the metric used to measure the proportion of true positive predictions out of all actual positive instances.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

- F1-score is a metric which combines both precision and recall into a single value to provide a balances assessment of a model's accuracy. It ranges between 0 and 1 and higher values indicate better model performance.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{3}$$

IV. Results & Discussion

The deep learning models were trained at kaggle online platform using T4 GPU of 15GB memory and 29GB CPU memory using 525 bird species datasets. After preprocessing the dataset, the class distribution shows 60% of them are resident species and 40% of them are migratory consisting features like body, color, beak, wings and more. The dataset includes mix of grassland, forest-dwelling, wetlands and urban adapted species, which suggests the presence of several threatened species, underscoring the importance of habitat conservation efforts.

Each convolutional layer will have few common terms which are found in the below subsections are *F(x)* which represents the outputs of the convolutional layers within residual block, x represents the input feature map, W and W_d are the set of learnable parameters(weights) of the convolutional layer and depthwise convolution respectively, *DW* represents depthwise convolution, *Swish* represents the Swish activation function, *BN* represents batch normalization, and *ReLU* is the rectified linear unit activation function

A. ResNet50

Residual Network model [8] with 50 layers, including convolutional, pooling and fully connected layers, can effectively leverage its depth to learn hierarchical features, resulting in improved performance on various computer vision tasks. It consists of several building blocks called residual blocks, and mathematical equation for a single residual block is expressed as eq. (4).

$$y = F(x) + x \tag{4}$$

After training, Fig. 3 can be observed and there is no large gap between training and validation losses and accuracies indicating model is in a spiky pace of obtaining generalization state (no overfitting or underfitting). ResNet50 model

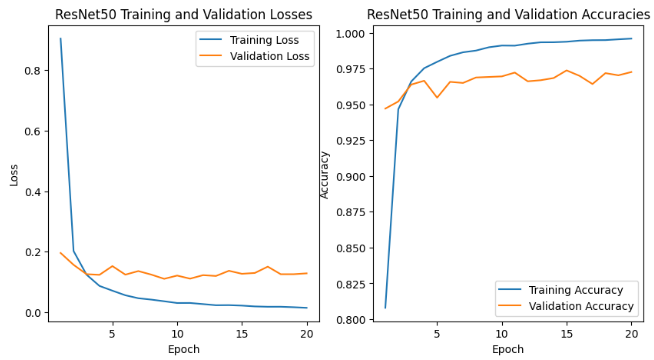

Fig. 3: ResNet50 Training and Validation Losses and Accuracies Plot

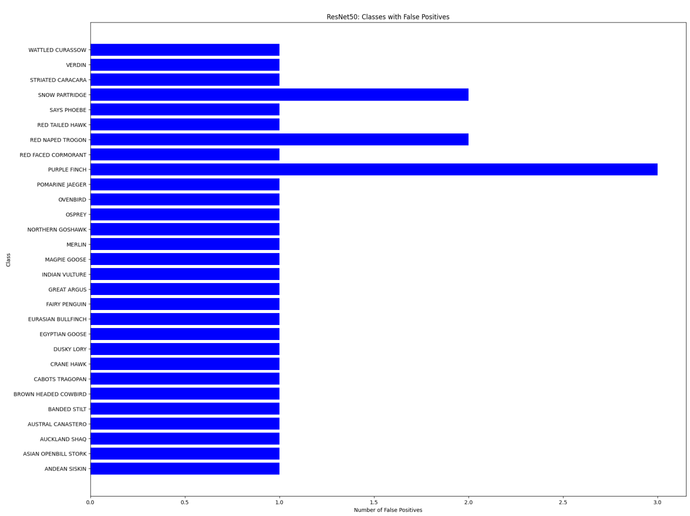

Fig. 4: ResNet50 Number of False Positives v/s Classes Plot

was tested, and obtained a score of 98.48% accuracy. Fig. 4 illustrates manual analysis where the model is predicting false positives for a set of species class. For 29 different bird species are observed to make false predictions due to similar color or body shape or eye color pattern features.

B. MobileNetV3 Small

MobileNetV3 belongs to a family of CNN architectures designed for mobile and edge devices with limited computational resources utilizing depthwise separable convolutions [9]. The mathematical equation for a single convolutional layer is expresses in the eq. (5).

$$y = ReLU(BN(DW(W_d * x))) \tag{5}$$

Fig. 5 can be observed where the model is in a steady pace of obtaining generalization state (no overfitting or underfitting). MobileNetV3 small model was tested, and obtained a score of 97.64% accuracy. Fig. 6 illustrates manual observation where the model made false predictions on 48 different species classes indicating high variance in

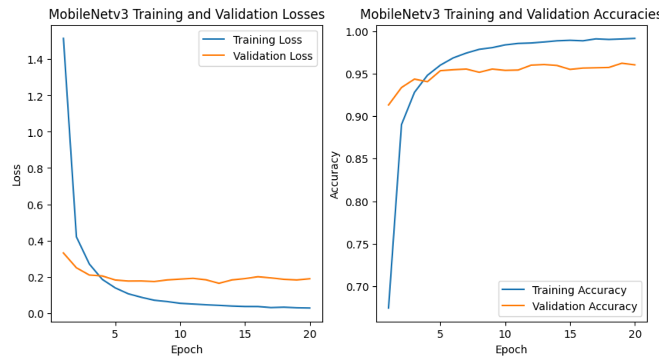

Fig. 5: MobileNetV3 Training and Validation Losses and Accuracies Plot

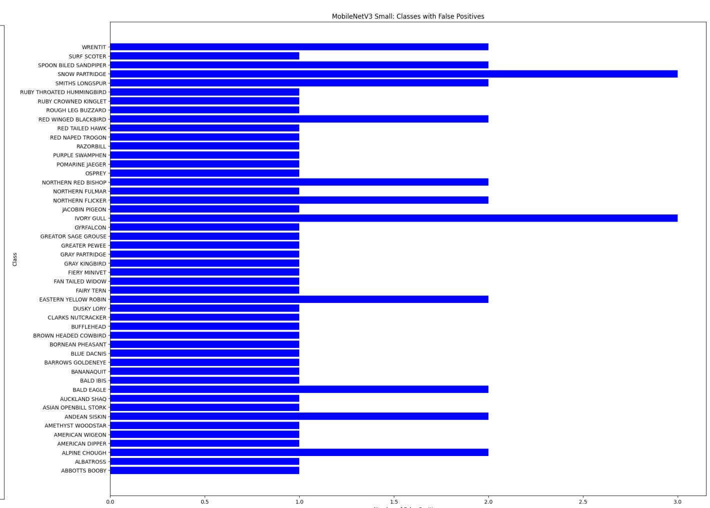

Fig. 6: MobileNetV3 Number of False Positives v/s Classes Plot

classifying them between different features, textures, shapes and with the background of images.

C. EfficientNet-B0

EfficientNet-B0 was designed to achieve state-of-the-art performance with significantly fewer parameters and computational resources when compared to traditional architectures [10]. The mathematical equation for a single convolutional layer is expressed as eq. (6).

$$y = Swish(BN(Conv(W * x))) \tag{6}$$

Fig. 7 illustrates the trained model having a small gap between train and validation loss/accuracy implying a good generalization state. This indicates the efficacy of model in achieving the intended goal. With the testing results, the model had 22 false prediction on different classes as shown in the Fig. 8 and on manually verification it is observed that these false predictions are caused by bird colors with different shades, lighting conditions and shape of them, even with those variations, it achieved a good accuracy of 99.09%.

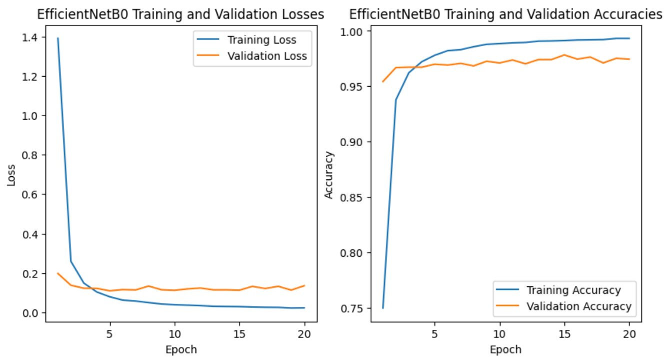

Fig. 7: EfficientNet-B0 Training and Validation Losses and Accuracies Plot

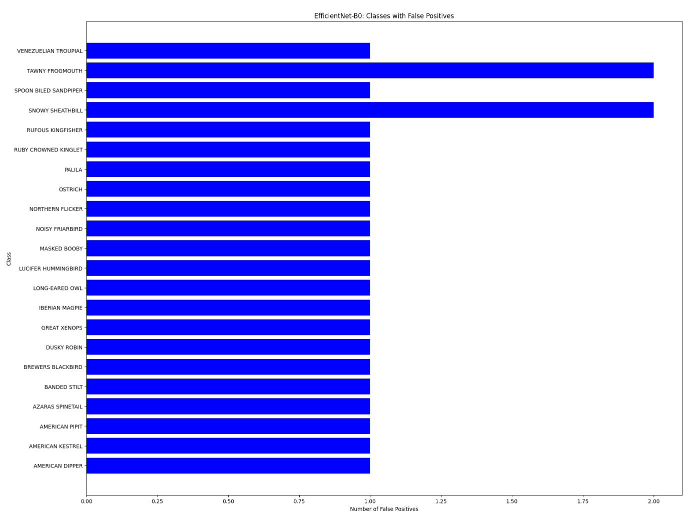

Fig. 8: EfficientNet-B0 Number of False Positives v/s Classes Plot

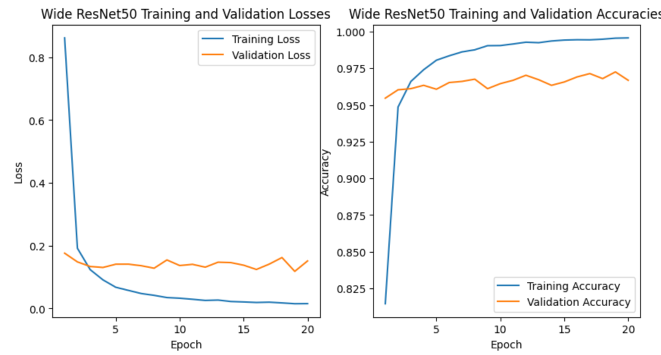

Fig. 9: Wide ResNet50V2 Training and Validation Losses and Accuracies Plot

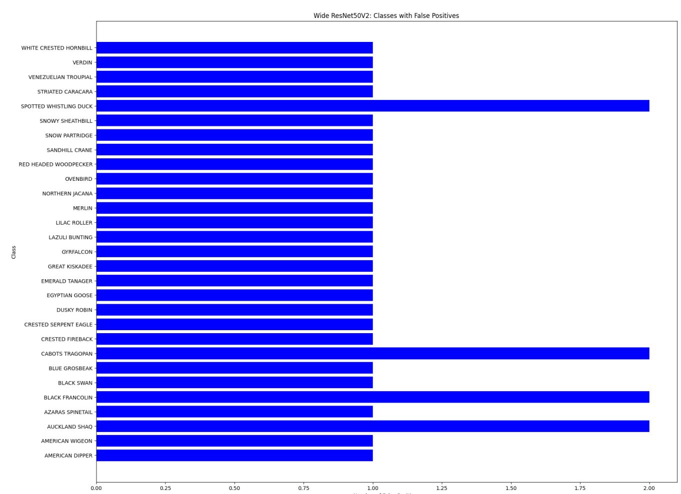

Fig. 10: Wide ResNet50V2 Number of False Positives v/s Classes Plot

D. Wide ResNet50V2

Wide ResNet50V2 is a deep convolutional neural network known for its wider and shallower design compared to traditional ResNet architectures. With its increased model width results in enhanced representational capacity and performance on various computer vision tasks, making it one of better choices for image analysis applications [11]. The mathematical equation for a single convolutional layer is expressed as eq. (7).

$$y = ReLU(BN(Conv(W * x))) \tag{7}$$

The graphs from the Fig. 9 are at a steady pace for both training and validation loss/accuracy and gaps have a good significant space with its wider design architecture. With the testing results, the model had 29 false prediction caused by bird colors with different shades, shapes, and lighting conditions on different classes as shown in Fig. 10, but it achieved a good accuracy of 98.74%.

With these analysis, this study compared four deep learning models-ResNet50, MobileNetV3 small, EfficientNet-B0, and Wide ResNet50V2 and assessed their performance on a 525 class bird species dataset. Every model had its own drawbacks and many characteristics playing a decisive role in causing false predictions on bird species, but none of the models had false negatives, that is, a positive class being incorrectly predicted as a negative class. Tab. 1 represents the performance metrics discussed in the evaluation metrics subsection, and apply those formulas to each of the CNN models based on their predictions with testing accuracy and validation loss of the best epoch.

The results demonstrate that EfficientNet-B0 architecture model has excelled with a precision of 99.09% and F1-score of 99.54% and is well suited for bird species classification tasks, outperforming the other CNN models in this study. This method achieves a classification accuracy of 99.09%, surpassing the one of the benchmarks of existing DenseNet169, EfficientNet-B0, ShuffleNetV2, average transfer learning ensemble approach model combining Efficient-

TABLE I: Model Evaluation Metric Table

Models	Testing Accuracy	Validation loss	Evaluation Metrics		
			Precision	***Recall***	***F1-score***
ResNet50	0.9848	0.1271	0.9848	1.00	0.9923
MobileNetV3	0.9764	0.1818	0.9764	1.00	0.9880
EfficientNetB0	0.9909	0.1127	0.9908	1.00	0.9954
Wide ResNet50V2	0.9874	0.1182	0.9874	1.00	0.9937

NetB5 and VGG19 and InceptionV3 on the 525 bird species testing dataset and these model analysis are compared in the Tab. 2 with testing accuracy and validation loss.

TABLE II: Comparitive analysis of Proposed System

Author and Citation	***Model***	***Testing Accuracy***	***Validation loss***
Proposed model	EfficientNet-B0	0.9909	0.1127
Chun Ing et al. [3]	DenseNet169	0.95	0.76
Cai Renjun [12]	EfficientNet-B0	0.867	0.49
Zhang et al. [13]	Modified ShuffleNetV2	0.911	0.41
Polishetty el al. [14]	Ensemble Model (EfficientNet-B5 and VGG19)	0.9718	0.1055
Shourie et al. [15]	InceptionV3	0.9866	1.0114

In summary, unlike previous methods that rely solely on feature engineering, this approach leverages advanced deep learning techniques to learn discriminating features from raw data. It allowed this model to achieve higher accuracy and robustness in bird classification tasks, showcasing the innovative potential of this proposed approach. An important note is that the performance of this model may vary, which will depend on the characteristics of the dataset and the complexity of the bird species. Images with different lighting conditions and angles, environment, varying background, poor quality of images, interclass variation, interspecies similarity, and more factors may decrease the performance of the model, which is one of the limitations of deep learning models even today. Models can be further trained with different hyperparameters or regularization techniques to improve performance and experiment with different learning rates, batch size, dropout rates, or changing any layer in the model architecture. Future research work will focus on overcoming the current limitation and fine-grain classification challenges to improve the performance and applicability of deep learning models in the field of ornithology.

V. Conclusion

Deep learning approach for bird species identification has yielded valuable insights into their applicability in real-world scenarios and performance. The EfficientNet-B0 model is effective in terms of computation resources and achieving high precision and F1-score on a 525 bird species training dataset comprising 84,635 images. The EfficientNet-B0 model is the best solution to identify different bird species, advancing the usage and necessity of deep learning technologies. However, there are a few limitations, such as different lighting conditions and angles, environmental factors, quality of images, inter-class variation, inter-species similarities, and more. The study contributes to the ongoing endeavour on model selection and paving the groundwork for future research. It also helps in biodiversity monitoring and species conservation.

References

[1] M. Y. Reddy, M. S. Ahmed, A. Nagumothu and M. Kavitha, "Analysis of DenseNet -MobileNet-CNN Models on Image Classification using Bird Species Data," 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, 2023, pp. 536-542, doi: 10.1109/ICDT57929.2023.10151357.

[2] T. Sonavane, K. Verma, and H. Wadile, "Automated Bird Species Identification Using Deep Learning With Image And Audio." EPRA International Journal of Multidisciplinary Research (IJMR), 8(11), 2022, 238-240.

[3] I. T. Chun Ing, N. S. Razali, M. A. A. Razak, E. S. L. Ming, N. A. Zakaria and N. H. Mahmood, "Deep Learning Analysis Using CNN For Bird Species Classification System," 2023 3rd International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA), Denpasar, Bali, Indonesia, 2023, pp. 107-111, doi: 10.1109/ICICyTA60173.2023.10428729.

[4] K. Hany, Y. Ayman and A. Atia, "Performance Evaluation of Pre-Trained Deep Learning Models for Bird Species Identification," 2023 Intelligent Methods, Systems, and Applications (IMSA), Giza, Egypt, 2023, pp. 309-314, doi: 10.1109/IMSA58542.2023.10255070.

[5] S. Neeli, C. S. R. Guruguri, A. R. A. Kammara, V. Annepu, K. Bagadi and V. R. R. Chirra, "Bird Species Detection Using CNN and EfficientNet-B0," 2023 International Conference on Next Generation Electronics (NEleX), Vellore, India, 2023, pp. 1-6, doi: 10.1109/NEleX59773.2023.10420966.

[6] Duta, P. K., and Nader Behdad. "A Deep Learning Approach Visual Recognition of Bird Species in Noisy Environments." Journal of Artificial Intelligence and Metaheuristics 3, no. 2 (2023): 18-8.

[7] TorchVision maintainers and contributors, "TorchVision: PyTorch's Computer Vision library," GitHub repository. [Online]. Available: https://github.com/pytorch/vision.

[8] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778, doi: 10.1109/CVPR.2016.90.

[9] A. Howard et al., "Searching for MobileNetV3," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019, pp. 1314-1324, doi: 10.1109/ICCV.2019.00140.

[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", arXiv:1905.11946 [cs.LG], 2020.

[11] S. Zagoruyko and N. Komodakis, "Wide Residual Networks", arXiv:1605.07146 [cs.CV], 2017.

[12] Cai, Renjun. "Automating bird species classification: A deep learning approach with CNNs." In Journal of Physics: Conference Series, vol. 2664, no. 1, p. 012007. IOP Publishing, 2023.

[13] Zhang, Liu-Lei, Ying Jiang, You-Peng Sun, Yuan Zhang, and Zheng Wang. "Improvements based on ShuffleNetV2 model for bird identification." IEEE Access (2023).

[14] Polisetty, Venkata Ramana Murthy, and S. P. Chokkalingam. "Enhancing Bird Species Classification: A Weighted Average Transfer Learning Ensemble Approach." In 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 1401-1406. IEEE, 2024.

[15] P. Shourie, V. Anand, R. Chauhan, D. Rawat, and S. Gupta, "A Dynamic System for Classifying Bird Species Using the InceptionV3 Model," in 2023 2nd International Conference on Futuristic Technologies (INCOFT), pp. 1-5, 2023.