Innovative Informatics and Artificial Intelligence Research (IIAIR)
Vol. 2, Issue 1, pp. 25–29, 2026
Received 10 Feb 2026; accepted 28 Apr 2026; published 30 Apr 2026
https://doi.org/10.35718/iiair.v2i1.8481939

Analysis of Lightweight Pretrained Deep Learning Models for Bird Species Classification

Remet Tirzah Anatasya[1] **and Rizal Kusuma Putra**[1]

[1]Department of Informatics, Institut Teknologi Kalimantan, Balikpapan, Indonesia

Corresponding author: Rizal Kusuma Putra (rizal.putra@lecturer.itk.ac.id)

To cite this article: R. T. Anatasya, R. K. Putra, "Analysis of Lightweight Pretrained Deep Learning Models for Bird Species Classification," *Innovative Informatics and Artificial Intelligence Research*, vol. 2, issue 1, 2026. [Online]. Available: https://doi.org/10.35718/iiair.v2i1.8481939

Abstract
Bird biodiversity plays a vital role in maintaining ecological balance but is highly vulnerable to urbanization and environmental degradation. Manual identification of bird species requires considerable time and expertise, making automated classification systems essential. This study develops an efficient bird species classification framework using lightweight deep learning models. Six pretrained architectures were evaluated: MobileViT-V1, MobileViT-V2, EfficientNetV2-B3, ResNet-18, MobileNetV3, and ShuffleNetV2. The dataset, obtained from TensorFlowDataset, consists of 200 bird species with a total of 11,788 images. Results indicate that EfficientNetV2-B3 achieved the highest accuracy (63%) and F1-score (0.63). MobileNetV3 provided a balanced trade-off with 56% accuracy and 0.835 ms latency which makes it suitable for resource-constrained device.

Keywords: bird species classification; deep learning; lightweight models; transfer learning; MobileViT; CNN; Vision Transformer

1. Introduction
Bird biodiversity is an important indicator of ecosystem balance and environmental health [1]. Birds play crucial ecological roles such as pollination, seed dispersal, and serving as bioindicators of environmental change [2]. Indonesia, as one of the world's megabiodiversity countries, possesses a very high richness of bird species [3]. The increasing number of species recorded each year emphasizes the importance of accurate identification and classification to support conservation, scientific research, and biodiversity education [1, 4].

However, manual identification of bird species remains challenging. It requires expert knowledge, significant time, and often faces difficulties in distinguishing morphologically similar species [5]. These limitations hinder large-scale monitoring and conservation efforts. Therefore, automated classification systems are needed to provide accurate, efficient, and consistent identification of bird species [6]. The advancement of computer vision and machine learning technologies offers promising solutions to overcome these challenges [7].

Deep learning, particularly Convolutional Neural Networks (CNN) and Vision Transformer (ViT) architectures, has demonstrated strong performance in image classification tasks [8, 9]. Transfer learning with pretrained lightweight models allows efficient adaptation to specific datasets, even when data is limited. In the context of resource-constrained devices, lightweight models are highly relevant due to their smaller parameter sizes, compact architectures, and low inference latency. This research aims to compare six lightweight pretrained models: MobileViT-V1 and MobileViT-V2, which integrate CNN efficiency with transformer-based feature learning; EfficientNetV2-B3 providing strong accuracy through optimized scaling; ResNet-18 serving as a widely adopted baseline with residual learning; MobileNetV3 which is specifically designed for mobile deployment with efficient latency-performance trade-offs; and ShuffleNetV2 emphasizing fast inference through channel shuffling and low computational cost. These models were selected to represent diverse lightweight design strategies for bird species classification and to provide recommendations for the most optimal model in terms of accuracy and computational efficiency [10, 11].

The contribution of this research lies in providing a systematic benchmark of lightweight pretrained models for bird species classification. Specifically, this study evaluates six architectures on a bird dataset (CUB-200-2011), analyzes the trade-off between accuracy and latency to highlight practical deployment considerations, and offers recommendations for selecting the most suitable model depending on application requirements. These contributions extend the application of transfer learning into biodiversity monitoring and conservation, supporting the development of efficient automated bird identification systems. The models chosen in this study MobileViT-V1, MobileViT-V2, EfficientNetV2-B3, ResNet-18, MobileNetV3, and ShuffleNetV2 were selected because they represent lightweight architectures widely used in computer vision research. They balance accuracy and compu-

tational efficiency, making them suitable for deployment on resource-constrained devices. Comparing these models provides insights into trade-offs between accuracy and latency, which is crucial for real-time applications.

2. Related Works

Previous studies have shown that deep learning, particularly convolutional neural networks (CNNs), is effective for image recognition tasks, including bird species classification [12, 10]. Architectures such as ResNet leverage residual connections to improve training stability and accuracy [13], while EfficientNetV2 applies compound scaling to achieve high accuracy with reduced computational cost [14]. Lightweight models such as MobileNetV3 and ShuffleNetV2 are widely used in mobile and embedded environments due to their low latency and compact architecture [15, 16]. In addition, hybrid models such as MobileViT integrate convolutional layers with transformer blocks to capture both local and global features, although this integration may increase inference latency compared to pure CNN-based models [17, 18].

Several studies have specifically applied deep learning to bird species classification. The study by Pillai et al. [19] presented bird classification research using transfer learning models. This study utilized 37,500 images divided into three sets: 24,000 training images, 7,500 test images, and 6,000 validation images, achieving a final accuracy of 97.12%. Huang et al. [10] developed a deep learning platform to assist users in recognizing 27 bird species endemic to Taiwan through a mobile application called Internet of Birds (IoB). Bird images were learned by a convolutional neural network (CNN) with skip connections, which achieved a higher accuracy of 99.00% compared to 93.98% from a standard CNN and 89.00% from SVM on the training images. For the test dataset, the average sensitivity, specificity, and accuracy were 93.79%, 96.11%, and 95.37%, respectively. Furthermore, a recent study by Zhang et al. [16] proposed improvements to the ShuffleNetV2 architecture for bird identification, which successfully enhanced computational efficiency while maintaining competitive accuracy . These findings indicate that although high accuracy can be achieved, considerations of computational efficiency and inference latency remain crucial to support deployment on resource-constrained devices.

Although existing studies report high classification accuracy, most of them primarily focus on predictive performance without explicitly considering computational efficiency and inference latency. In contrast, this study conducts a comparative evaluation of six lightweight pretrained models for bird species classification, emphasizing the trade-off between accuracy and latency to support deployment on resource-constrained devices.

3. Methods

This study employed a comparative experimental approach using lightweight deep learning models to classify bird species automatically. The models selected for comparison were MobileViT-V1, MobileViT-V2, EfficientNetV2-B3, ResNet-18, MobileNetV3, and ShuffleNetV2. These architectures were chosen because they represent lightweight designs with relatively small parameter counts (1–15 million), low latency, and availability in popular deep learning libraries such as TensorFlow and PyTorch. Transfer learning was applied to adapt pretrained weights to the bird dataset, enabling efficient training despite limited data.

3.1. MobileViT-V1

MobileViT-V1 is a hybrid architecture that integrates Convolutional Neural Networks (CNN) and Vision Transformers to efficiently learn both local and global features [17]. Local features are first extracted using convolution operations as shown in Equation 1

$$F_conv = W * X + b \tag{1}$$

The resulting feature maps are then partitioned into small patches and processed using a self-attention mechanism to capture global dependencies as shown in Equation 2

$$Attention(Q, K, V) = Softmax(\frac{QK^T}{\sqrt{d_k}}) \tag{2}$$

The transformer output is reshaped back into a spatial feature map and fused with convolutional features, preserving the spatial inductive bias of CNN while reducing the computational cost compared to standard Vision Transformer models.

3.2. MobileViT-V2

MobileViT-V2 is an improved version of MobileViT-V1 that introduces optimizations in the transformer blocks and feature fusion mechanisms. While the core principle of combining convolutional local feature extraction and transformer-based global modeling remains unchanged, the refined architecture reduces parameter redundancy and inference latency. As a result, MobileViT-V2 achieves a better accuracy–efficiency trade-off and is more suitable for deployment on resource-constrained devices [20].

In this study, pretrained MobileViT-V2 was employed using a transfer learning strategy, where the backbone layers were frozen and only the final classification head was fine-tuned for bird species classification.

3.3. EfficientNetV2-B3

EfficientNetV2-B3 is a variant of the EfficientNetV2 family designed to achieve an optimal balance between classification accuracy and computational efficiency for image recognition tasks [19]. The model adopts the compound scaling strategy, in which network depth, width, and input resolution are jointly scaled to improve performance without incurring a significant increase in computational cost [14]. The EfficientNetV2-B3 architecture is composed of a combination of Fused-MBConv and MBConv blocks, which are derived from depthwise separable convolutions and enable parameter reduction while preserving rich feature representations.

Compared to smaller EfficientNet variants, EfficientNetV2-B3 employs increased network depth and higher input resolution, allowing the model to capture more complex visual patterns from large-scale datasets [18]. In addition, the use of the Swish activation function contributes to improved training stability and enhanced model accuracy. In this study, EfficientNetV2-B3 is employed as a pretrained backbone using a transfer learning approach, where the backbone layers are frozen and only the final classification head is fine-tuned for the bird species classification task.

Table 1: Comparative Performance across Lightweight Models

Models	Accuracy	Precision	Recall	F1-score	Latency
MobileViTV1	0.42	0.46	0.43	0.42	2.514 ms
MobileViTV2	0.57	0.59	0.57	0.56	1.833 ms
EfficientNetV2-B3	**0.63**	**0.66**	**0.64**	**0.63**	1.753 ms
ResNet-18	0.53	0.57	0.54	0.53	0.898 ms
MobileNetV3	0.56	0.60	0.56	0.55	0.835 ms
ShuffleNetV2	0.34	0.39	0.35	0.32	**0.220** ms

3.4. ResNet-18

ResNet-18 is a convolutional neural network belonging to the Residual Network (ResNet) family, originally proposed by He et al. to address the vanishing gradient problem in deep neural networks [13]. The key component of ResNet-18 is the use of residual connections, which enable the network to learn residual mappings and facilitate efficient gradient propagation during training [21]. A residual block can be expressed in Equation 3

$$y = F(x) + x \tag{3}$$

where F(x) represents the nonlinear transformation of the input feature x. This design allows information to bypass several convolutional layers, improving training stability and convergence. Due to its relatively shallow depth and moderate computational complexity, ResNet-18 is commonly used as a baseline architecture in image classification tasks. In this study, ResNet-18 is employed as a pretrained model, with the backbone layers frozen and only the final classification layer fine-tuned for bird species classification.

3.5. MobileNetV3

MobileNetV3 is a lightweight convolutional neural network designed for deployment on resource-constrained devices such as mobile phones and Internet-of-Things (IoT) platforms. The architecture aims to achieve a balance between computational efficiency and classification accuracy by incorporating optimized building blocks derived from both manual design and Neural Architecture Search (NAS) [22].

MobileNetV3 employs depthwise separable convolutions and bottleneck blocks to significantly reduce the number of parameters and floating-point operations. In addition, the architecture integrates Squeeze-and-Excitation (SE) modules to enhance channel-wise feature representations with minimal computational overhead. A key innovation of MobileNetV3 is the use of the hard-swish activation function, which improves inference efficiency while maintaining competitive performance compared to ReLU-based activations [15].

Due to its compact architecture and low inference latency, MobileNetV3 has been widely adopted in image classification, object detection, and segmentation tasks. In this study, MobileNetV3 is utilized as a pretrained backbone, where the feature extraction layers are frozen and only the final classification head is fine-tuned for bird species classification.

3.6. ShuffleNetV2

ShuffleNetV2 is a lightweight convolutional neural network designed to achieve high computational efficiency, particularly on resource-constrained platforms such as mobile and Internet-of-Things (IoT) devices. The architecture was introduced to address the practical runtime limitations of earlier lightweight models by explicitly optimizing memory access cost and computational balance [16]. The core design of ShuffleNetV2 follows a channel split-and-shuffle strategy, which enables efficient information exchange across feature channels while minimizing computational overhead. This operation improves feature utilization without increasing model complexity.

In this study, ShuffleNetV2 is employed as a lightweight pretrained baseline model. The backbone layers are frozen, and only the final classification layer is fine-tuned for bird species classification, allowing an effective evaluation of the trade-off between classification accuracy and inference latency.

4. Experimental Settings

This study adopts a comparative experimental approach to evaluate the performance of several lightweight deep learning architectures for automatic bird species classification. All models were trained using a transfer learning strategy with pretrained weights from ImageNet. The models were trained for 50 epochs using the AdamW optimizer and the CrossEntropy loss function. The learning rate was set to 0.001, with a batch size of 32 and a weight decay of 1×10^{-4}. Model performance was evaluated using accuracy, precision, recall, F1-score, and latency metrics. Latency was measured to assess suitability for real-time applications in GPU device.

In this study, each pre-trained architecture was executed using a transfer learning strategy. Pre-trained weights from ImageNet were employed, and the majority of the backbone layers were frozen to preserve general feature representations and accelerate training. Only the final classifier head was fine-tuned to adapt to the bird dataset (200 species) [23]. This approach reduces computational cost, prevents overfitting on a relatively limited dataset, and ensures faster inference.

5. Results and Discussios

The comparative performance results are summarized in Table 1. EfficientNetV2-B3 achieved the best overall performance, obtaining the highest accuracy, precision, recall, and F1-score among all evaluated models, while ShuffleNetV2 exhibited the fastest inference time. EfficientNetV2-B3 delivered approximately 30% higher accuracy than ShuffleNetV2 at the cost of an additional 1.5 ms in inference time. However, despite its computational efficiency, ShuffleNetV2 achieved only around 34% accuracy and 32% F1-score, indicating a substantial trade-off between speed and predictive performance. although EfficientNetV2-B3 achieved the highest accuracy of 63%, the result reflects the difficulty of distinguishing a large number of visually similar categories rather than poor model capability. In contrast, ShuffleNetV2 obtained

only 34% accuracy because its highly compact architecture prioritizes computational efficiency and low latency, which may limit its feature extraction capacity for complex fine-grained classification involving 200 bird species.

The baseline ResNet-18, which leverages residual connections, remains competitive compared to other architectures. Although its accuracy is approximately 10% lower than that of EfficientNetV2-B3, ResNet-18 outperformed both MobileViT-V1 and ShuffleNetV2. In terms of computational efficiency, ResNet-18 achieved an inference time below 1 ms, making it faster than EfficientNetV2-B3 and all MobileViT variants.

In contrast, MobileViT-V1 and MobileViT-V2 achieved only moderate accuracy while suffering from higher inference latency due to their hybrid CNN–Transformer architectures. MobileViT-V1, which was expected to outperform the ResNet-18 baseline, failed to do so and exhibited the slowest inference time, requiring approximately 2.5 ms per prediction. MobileViT-V2 partially alleviated this limitation by improving accuracy to approximately 57%, making it the second-best performing model and about 1% higher than MobileNetV3.

Figure 1 illustrates the trade-off between model performance and computational efficiency by plotting F1 Score against latency for each of the six evaluated architectures. EfficientNetV2-B3 is positioned at the top-right quadrant, indicating the highest F1 Score (0.63) but with relatively high latency (1.8 ms), making it ideal for accuracy-focused applications but less suitable for real-time deployment.

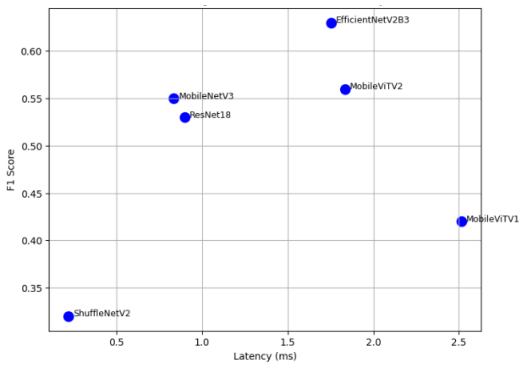

Figure 1: F1-score vs Latency for all comparison models

MobileNetV3 stands out in the middle-left region, offering a strong F1 Score (0.55) with low latency (0.84 ms), which confirms its balanced nature and suitability for resource-constrained environments. ResNet-18, while slightly lower in F1 Score (0.53), maintains competitive latency (0.898 ms), showing that residual connections remain effective in lightweight scenarios.

ShuffleNetV2 appears at the bottom-left corner, achieving the lowest latency (0.2 ms) but also the lowest F1 Score (0.34), indicating its strength in speed but weakness in predictive accuracy. MobileViT-V1 and MobileViT-V2 are located toward the right side of the plot, reflecting higher latency due to their hybrid CNN-Transformer architecture, with MobileViT-V2 showing improved F1 Score (0.56) compared to V1 (0.43). Overall, the scatter plot in Figure 1 highlights the importance of selecting models based on application requirements. EfficientNetV2-B3 is optimal for high-accuracy tasks, MobileNetV3 offers the best balance, and ShuffleNetV2 is preferable when ultra-low latency.

The results emphasize the importance of selecting models based on application requirements. EfficientNetV2-B3 is optimal when accuracy is prioritized, MobileNetV3 offers the best balance for practical deployment, and ShuffleNetV2 may be considered in scenarios where ultra-low latency is critical but accuracy requirements are minimal.

6. Conclusion

This study compared six lightweight pretrained models for bird species classification. The findings revealed that EfficientNetV2-B3 achieved the highest accuracy and F1-score, ShuffleNetV2 provided the lowest latency but poor accuracy, and MobileNetV3 demonstrated strong competitive performance while offering the most balanced trade-off between accuracy and computational efficiency.

In conclusion, MobileNetV3 is recommended as the optimal model for deployment on resource-constrained devices due to its ability to maintain reliable classification performance with efficient computational cost, highlighting its success as a practical solution for real-world bird species identification. Meanwhile, EfficientNetV2-B3 is suitable for applications requiring higher accuracy. These results contribute to the development of efficient automated bird identification systems and have practical implications for biodiversity monitoring and bird conservation in Indonesia, where efficient field-deployable systems can support species observation, habitat assessment, and conservation decision-making in remote areas. Furthermore, the findings emphasize the importance of balancing predictive performance with computational limitations. Future research should investigate hybrid CNN-transformer architectures, model compression techniques such as pruning and quantization, real-time mobile deployment in field environments, and the use of larger or more diverse Indonesian bird datasets. In addition, future studies should provide a more comprehensive efficiency analysis by evaluating FLOPs, the number of parameters, and memory usage to better assess model suitability for practical deployment.

References

[1] T. Lamba, H. H. Pontororing, and Saroyo, "Biodiversitas burung pada beberapa tipe habitat di kampus universitas sam ratulangi manado dalam masa pandemi covid-19," *JURNAL LPPM BIDANG SAINS DAN TEKNOLOGI*, vol. 7, no. 2, p. 27–32, Oct. 2022, doi: 10.35801/jlppmsains.7.2.2022.47488 . [Online]. Available: https://ejournal.unsrat.ac.id/v3/index.php/lppmsains/article/view/47488

[2] R. Fabrina and U. Faizah, "Keanekaragaman dan kelimpahan jenis burung di kawasan mangrove bee jay bakau resort (BJBR) kota probolinggo," *Sains & Mat*, vol. 7, no. 1, pp. 1–7, Apr. 2022, doi: 10.26740/sainsmat.v7n1.p1-7.

[3] https://burung.org/author/burung admin/, "Status Burung di Indonesia 2024 — burung.org," https://www.burung.org/en/status-burung-di-indonesia-2024/, [Accessed 19-11-2024].

[4] H.-T. Vo, N. N. Thien, and K. C. Mui, "Bird detection and species classification: Using yolov5 and deep transfer learning models," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 7, 2023, doi: 10.14569/IJACSA.2023.01407102. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2023.01407102

[5] D. F. Lestari and I. Kurnia, "Keanekaragaman jenis burung pada berbagai tipe habitat di pulau belitung," *Bioscientist : Jurnal Ilmiah Biologi*, vol. 11, no. 1, p. 1–19, Jun. 2023, doi: 10.33394/bioscientist.v11i1.6725. [Online]. Available: https://e-journal.undikma.ac.id/index.php/bioscientist/article/view/6725

[6] A. H. Abdel-aziem and T. H. M. Soliman, "A Multi-Layer Perceptron (MLP) Neural Networks for Stellar Classification: A Review of Methods and Results," *International Journal of Advances in Applied Computational Intelligence*, no. Issue 2, pp. 29–37, Jan. 2023, doi: 10.54216/IJAACI.030203. [Online]. Available: https://www.americaspg.com/articleinfo/31/show/2002

[7] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, 2023, doi: 10.3390/computation11030052. [Online]. Available: https://www.mdpi.com/2079-3197/11/3/52

[8] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadhel, M. Al-Amidie, and L. Farhan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.

[9] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, Z. Yan, M. Tomizuka, J. Gonzalez, K. Keutzer, and P. Vajda, "Visual transformers: Token-based image representation and processing for computer vision," 2020, doi: 10.48550/arXiv.2006.03677. [Online]. Available: https://arxiv.org/abs/2006.03677

[10] Y.-P. Huang and H. Basanta, "Bird image retrieval and recognition using a deep learning platform," *IEEE Access*, vol. 7, pp. 66 980–66 989, 2019, doi: 10.1109/ACCESS.2019.2918274.

[11] X. Zhao, L. Wang, Y. Zhang, X. Han, M. Deveci, and M. Parmar, "A review of convolutional neural networks in computer vision," *Artif. Intell. Rev.*, vol. 57, no. 4, Mar. 2024, doi: 10.1007/s10462-024-10721-6.

[12] Z. Wang, J. Wang, C. Lin, Y. Han, Z. Wang, and L. Ji, "Identifying habitat elements from bird images using deep convolutional neural networks," *Animals*, vol. 11, no. 5, 2021, doi: 10.3390/ani11051263. [Online]. Available: https://www.mdpi.com/2076-2615/11/5/1263

[13] G. K. Pandey and S. Srivastava, "Resnet-18 comparative analysis of various activation functions for image classification," in *2023 International Conference on Inventive Computation Technologies (ICICT)*, 2023, pp. 595–601, doi: 10.1109/ICICT57646.2023.10134464.

[14] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97. PMLR, 09–15 Jun 2019, pp. 6105–6114. [Online]. Available: https://proceedings.mlr.press/v97/tan19a.html

[15] S. Qian, C. Ning, and Y. Hu, "Mobilenetv3 for image classification," in *2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE)*, 2021, pp. 490–497, doi: 10.1109/ICBAIE52039.2021.9389905.

[16] L.-L. Zhang, Y. Jiang, Y.-P. Sun, Y. Zhang, and Z. Wang, "Improvements based on shufflenetv2 model for bird identification," *IEEE Access*, vol. 11, pp. 101 823–101 832, 2023, doi: 10.1109/ACCESS.2023.3314676.

[17] S. Mehta and M. Rastegari, "Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer," 2022, doi: 10.48550/arXiv.2110.02178. [Online]. Available: https://arxiv.org/abs/2110.02178

[18] V. R. M. Polisetty and S. Chokkalingam, "Efficient classification of bird species using photographic images: A mobilevit based approach," in *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*, 2024, pp. 1–6, doi: 10.1109/AIIoT58432.2024.10574683.

[19] R. Pillai, N. Sharma, D. Upadhyay, S. Devliyal, and G. Kaur, "Efficientnet-v2-b3 transfer learning for high-accuracy indian bird species classification," in *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2024, pp. 1–6, doi: 10.1109/ICCCNT61001.2024.10724769.

[20] S. Mehta and M. Rastegari, "Separable self-attention for mobile vision transformers," 2022, doi: 10.48550/arXiv.2206.02680. [Online]. Available: https://arxiv.org/abs/2206.02680

[21] S. A. Al-Showarah and S. T. Al-qbailat, "Birds identification system using deep learning," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, 2021, doi: 10.14569/IJACSA.2021.0120434. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2021.0120434

[22] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, Q. V. Le, and H. Adam, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[23] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

