Grenze International Journal of Engineering and Technology, January Issue

Prediction of Endangered species using variants of EfficientNet

Anamika Thakur[1], Pari Gupta[2], Sakshi Mittal[3] and Shweta Jindal[4]
[1-4]Indira Gandhi Delhi Technical University for Women, Delhi, India
Email: anamika2642005@gmail.com, pariguptaa3@ gmail.com, sakshimittal11350@gmail.com, miss.shweta.singhal@gmail.com

***Abstract*— Biodiversity conservation is critical in addressing the growing threats to endangered species worldwide. Accurate prediction of endangered species status is vital for developing targeted conservation strategies. This study explores the application of machine learning models, specifically the EfficientNetV2 family of architectures, to predict endangered species using image data. The training process employed a two-stage approach. Initially, the pre-trained EfficientNet models (variants B0, B1, B2, B3, S, M, and L) were fine-tuned by freezing their base layers and training custom classification heads using sparse categorical cross-entropy loss. Subsequently, the entire models were unfrozen for comprehensive fine-tuning at reduced learning rates, optimizing feature extraction for the specific task. The results highlight the potential of advanced machine learning architectures for biodiversity conservation, offering a scalable and accurate solution for species identification and classification. This study emphasizes the importance of integrating computational methods with conservation efforts to address global challenges in preserving biodiversity.**

***Index Terms*— Endangered, Biodiversity, Efficient Net, Image Processing.**

I. INTRODUCTION

As habitat deterioration increases, the impacts of climate change on a macro level and the loss of biodiversity, the goals of wildlife conservation become increasingly complex and essential. It is timely to note that it has profound impacts that go beyond maintaining the sensitive balance of ecosystems or merely protecting the wellbeing of the Earth. Nowadays, forests are crucial to mitigate the consequences of global warming, as well as control the water quality and serve as habitats for different wildlife species. Nowadays, forests assume a critical role, due to their capability to mitigate global warming. They capture carbon dioxide, enhance the water quality, moderate city temperature, and serve as home to many wildlife species. The methods of tree planting and tree cutting are largely manual and therefore, expensive, time consuming and unsafe. Still, due to the manual nature of most forestry work, costs are high and those who do the work are placed in potentially harmful situations.
Forestry includes an array of activities relating to forest monitoring, inventory, management, wildlife monitoring, search and rescue, and the monitoring of wildfire. With all these endeavors, extensive data are collected and combined concerning the forests, such as the height and diameter of trees, species composition, and population density, thus giving insights on what happened to the forests until now and, in turn, being used as inputs for forest management.

Grenze ID: 01.GIJET.12.1.25

The main thrust of forest management is the protection and maintenance of timbered areas, responding to monitoring data that in turn implies actions such as removal of diseased vegetation or reforestation. Wildlife monitoring is concerned with monitoring different types of wildlife inhabiting these forests. Search and rescue and wildfire monitoring are built-in to the aspect of crisis management, which is to find people in need of assistance and to limit the damage in cases of emergency in forests.

The integration of drones and LiDAR [1] , has brought forth a revolution in the modes of data collection in forestry and wildlife monitoring. These aerial platforms provide researchers and conservationists a vantage point to survey wildlife habitats, which, otherwise, would be very hard to access. With such precise surveys and tracking of species conducted in remote areas, the ever-evolving machine learning and computer vision algorithms, combined with the drones, hold various advantages including precise and effective identification as well as monitoring of species for wildlife protection. Such technological advancement has the potential to reshape conservation efforts worldwide.

The use of drones has revolutionized the collection of data from challenging habitats and remote environments. With high-resolution cameras on drones, researchers and specialists dealing with wildlife can gain proper knowledge about wildlife habitats and soil characteristics with a view that otherwise would not have been possible to reach [2]. In fact, the enormous amount of data produced from drone-built images constitute a huge hurdle, that motivates the application of modern deep learning algorithms, particularly CNN. The automatic learning approach functions well for feature extraction in images, making them highly appropriate for tasks involving classification and recognition of wildlife species from drone-scanned images. The various applications of this transformative technology can definitely automate the classification and recognition of different wildlife species and can, after all, contribute to the protection of endangered species [3].

The primary goal in this consideration is to broaden the array of tools available to conservationists so that they are well-equipped to make wise decisions and lead proactive initiatives dedicated to the protection of Earth's extraordinary and vulnerable species. We fully embrace these paradigm-making technologies as we join our forces to preserve the future of our planet's rich and irreplaceable wildlife [3].

II. Literature Review

Gray's [5] paper explains the difficulty of observing marine megafauna because of their migration patterns and sparse population data, making conservation difficult. Drones offer a possible solution, although data analysis may be challenging. Neural networks can be used to automate object detection in drone images, improving understanding of marine populations. In a study of olive ridley turtles in Costa Rica, neural networks detected 8% more turtles than visual counts and improved validation. Successful training of the model on a limited dataset reveals its feasibility for other applications, efficiently mixing drone technology with neural networks for more efficient monitoring of marine life.

Rominger's [6] research highlights the need for proper census data in the management of endangered flora, while pointing out that resource constraints can impede thorough on-the-ground censuses. Accessibility is also a concern for census activities, especially for species in sensitive habitats. The research investigated the potential use of UAV (drone) technology to census the endangered plant Arctomecon humilis (dwarf bear-poppy) in the Mojave Desert, USA. Imagery was taken at 50-m and 15-m altitudes over two populations, yielding high-resolution orthomosaics. Although the 50-m imagery served for preliminary plant identification, follow-up verification using the 15-m imagery identified errors in both detection and recognition of smaller plants, emphasizing precision difficulties.

Kellenberger's [7] article elaborates on the application of semi-automatic models based on UAV images to aid in counting huge mammals in large territories such as the African Savanna. The scarcity of animals in relation to the large number of images leads to high false positives in detection. The study proposes a methodology based on a CNN and presents several recommendations to improve model training, including class-weighting, curriculum learning, hard negative mining, and introducing a border class for better detection accuracy. An alternative evaluation protocol is introduced that focuses on the number of detections rather than precise localization, as well as a second method for assessing detection efficiency in relevant image tiles. The effectiveness of these approaches was demonstrated in the Kuzikus game reserve in Namibia, where the CNN outperformed a state-of-the-art detector in precision and reduced the number of tiles requiring manual verification by over two-thirds at a 90% recall level.

Hua's [8] study focuses on the challenges of monitoring at-risk megafauna, specifically the black rhino, and presents a drone-based model for automated detection. Traditional monitoring methods are often invasive and costly, while satellite monitoring lacks the resolution needed for smaller species. The proposed system utilizes a

Jetson Xavier NX onboard a Parrot Anafi drone, allowing for lightweight, web-based notifications when target species are detected, even in areas with sporadic internet connectivity through MQTT messaging. Using a YOLOv5l6 object detection model, the model recorded an average precision of 0.81 for black rhinos and 0.83 for giraffes, although it performed lower for smaller animals. The model was trained on a varied dataset to reduce false positives and improve performance across different field conditions, showing the potential for automated wildlife detection and alerting in remote areas.

Wich et al. [9] give a detailed summary of the uses of drones in conservation. Their discussion points out how, with specific conditions and proper planning, drones can be used in a way that causes minimal disturbance to wildlife so that these sophisticated tools serve to improve conservation efforts without adversely affecting animal behavior or habitats.

Zhang et. al. [10], adapted the YOLOv8 and YOLOv3 architectures, for aerial imagery of urban environments. They primarily used the VisDrone2019 dataset.

For instance, in [11], a faster R-CNN model was developed to detect poachers and animals in drone-captured videos. The study provides a comparison of inference times when processing these videos on a CPU, GPU, and cloud platform. The Faster R-CNN model results in an inference rate of approximately 5 frames per second with K40 GPU alone, according to key findings. However, as live video streams typically run at 25 fps, this discrepancy causes a bottleneck in the detection process, leading to challenges with video synchronization.

In the research study by Xu et. al. [12] , The drone was flown at an altitude of 60–70 meters, and using the YOLOv3 model, it achieved an accuracy of 86.7% mean Average Precision (mAP) with a frame rate of 17 fps. In comparison, YOLOv2 reached a lower accuracy of 49.3% mAP but processed at a faster rate of 52 fps. However, it is difficult to make out how the model performed as the results are not discussed.

In [13] , training was done on a limited dataset (350 images per class). Improvement that should be done is to train the model on a large dataset and then improve the accuracy.

In [14], it use the CNN model along with SVM and they used the self made data to achieve a high accuracy of 96.02%. However the study broadly classifies into mammal, bird, fish, reptile, amphibia and not much significant in estimating endangered species in particular.

III. METHODOLOGY

A. Introduction to Efficient Net Architecture

EfficientNet is a family of CNNs created for the purposes of maximizing either accuracy and efficiency across various model scales. EfficientNet was introduced by Tan and Le as an approach that used a method of compound scaling[15]. This is different from previous approaches to scaling networks that tended to adjust depth, width, and resolution separately, which is not always optimal. EfficientNet-B0 is the baseline model that is found after a Neural Architecture Search (NAS) in the MobileNetV2 architecture space, and the additional larger scales, EfficientNet-B1 to B7, are built from scale penalties applied to the baseline.

The baseline model EfficientNet-B0 has a variety of MBConv blocks; MBConv stands for Mobile Inverted Bottleneck Convolutions initiated in MobileNetV2 [16]. Each block is characterized by an expansion, a depthwise separable convolution, and a projection of that downsampled layer; most blocks also have a squeeze-and-excitation (SE) optimization, which provides a balance between performance and cost in the computational budget.[17]

The efficient architecture begins with a normal convolution and max pooling layer, and moves onto many MBConv blocks based on our starting size with a gradually adusting sized filter to downsample the spatial resolution. A notable understanding about EfficientNet is the switch activation function (a smooth, non-monotonic function) that is used throughout the network, which has been shown to optimally outperform the classic ReLU activation function [18].

Rather than just uniformly scaling dimensions without consideration, EfficientNet employs a compound coefficient ϕ (phi) to scale:

- Depth (d): Number of layers (controls representational power)
- Width (w): Number of channels (controls capacity)
- Resolution (r): Input image size (controls spatial detail)

For a nominal set of scaling factors α, β, γ and identified through a grid search, the scale network dimensions are:

- Depth = $\alpha\phi$,
- Width = $\beta\phi$,
- Resolution = $\gamma\phi$

subject to the constraint:

- $\alpha \cdot \beta 2 \cdot \gamma 2 \approx 2$

This compound approach provides a balanced model scaling and avoids choosing a network that is too large (overfitting) or too small (under-utilization).

TABLE II: SCALING FACTORS AND MODEL SIZES FOR EFFICIENTNET VARIANTS

Model	φ	Depth	Width	Resolution	Parameters (M)	Top-1 Accuracy (%)
EfficientNet-B0	0	1.0x	1.0x	224x224	5.3	77.1
EfficientNet-B1	1	1.1x	1.0x	240x240	7.8	79.1
EfficientNet-B4	4	1.4x	1.8x	380x380	19.0	83.0
EfficientNet-B7	7	2.0x	2.0x	600x600	66.0	84.4

After the relative success of the original EfficientNet models, various variants were developed:

- EfficientNet-L2: A larger variant touted for transfer learning, scaled beyond B7 using noisy student training technique.
- EfficientNetV2: Improves training speed and parameter efficiency by mixing fused-MBConv blocks (standard convolution in place of depthwise convolution in the block's earliest layers) with progressive training strategies.
- EdgeTPU & Lite variants: Designed for deployment on mobile and edge devices with optimized latency and memory footprint.

EfficientNet models attained new cutting-edge performance on ImageNet with significantly smaller parameter counts and FLOPs than prior architectures. EfficientNet-B4, for example, achieves higher accuracy with nearly 10x fewer parameters than ResNet-152.

The architecture's scalability, principled design, and hardware-awareness make it a strong baseline and transfer-learing backbone for a wide range of vision tasks—from classification and detection to segmentation and more.

B. Data Preprocessing

To ensure the dataset was prepared for effective machine learning model training, several preprocessing steps were implemented. The dataset utilized for this study, "Animal Image Dataset (90 Different Animals)," was obtained from Kaggle and consisted of 5,400 images classified into 90 unique animal classes. The dataset was divided into three subsets: 70% for training (3,780 images), 15% for validation (810 images), and 15% for testing (810 images)[19].

Firstly all images are resized to 256x256 pixels, as the efficient netV2 -B1 model requires fixed input dimensions. Then pixel values are normalized by scaling them in the range [0,1] for model training.

Then, class labels were one-hot encoded, it is an approach for multi-class classification tasks, to use categorical cross-entropy as the loss function during training. Finally we prepared the dataset in mini batches where the batch size was kept as 32. All these preprocessing steps were performed to accurately predict endangered species.

C. Model Architecture

The model is trained using Google Colab which provided access to an NVIDIA T4 GPU. This helped us to speed up the training process compared to using a regular computer. The methods presented in this paper are several versions of EfficientNetV2 named B0, B1, B2, B3, S, M, and L, differentiated on the basis of capacity and size. The study has used the variants of EfficientNet for the prediction of endangered species which differ in their memory usage, speed and performance.

Our model has compound scaling and is used for scaling network depth and width using scaling coefficients. Our model also has Fused-MBConv layers. Fused-MBConv layers are an improved version of the original Mobile Inverted Bottleneck Convolution layers used in earlier EfficientNet models. These layers streamline the process by merging depthwise convolutions and expansion operations, making training and inference significantly faster while lowering computational costs. This optimization not only enhances efficiency but also improves overall performance, which made it a valuable innovation for models like ours, especially on resource-constrained devices.

The custom classification head was made up of three key parts designed to work together seamlessly for effective multi-class classification:

Global Average Pooling: This layer takes the feature maps from the base model and reduces their size by averaging the values in each feature map. It creates a single, compact feature vector for each channel, making the data easier to process in the next layers.

Dropout: To prevent the model from overfitting — where it gets too good at memorizing the training data but struggles with new data — a dropout layer was added with a rate of 0.2. This means that during training, it randomly "turns off" some neurons, which encourages the model to learn more robust features and rely on different parts of the network.

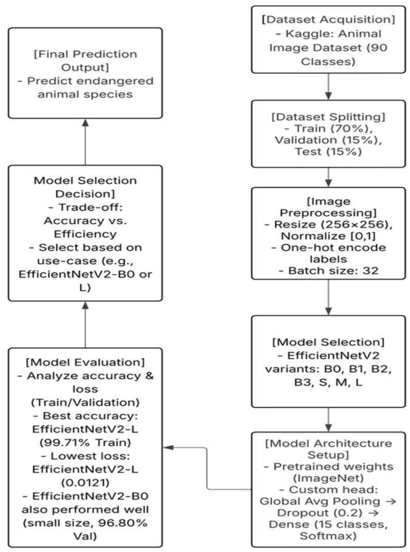

Figure.1 Whole process in a nutshell

Dense Layer: Finally, the dense layer outputs the class probabilities, with its size matching the number of classes (15, here). Using a softmax activation function, it turns the model's final output into a set of probabilities for each class, making it easier to determine which class the input most likely belongs to.
EfficientNetV2-B1 is the best model for balancing computational requirements we had and the accuracy we got. It uses image net weights for transfer learning which is the starting point for feature extraction.
Although the architecture of the base model is same for all the variants of EfficientNetV2 variants used that is NetV2, B0, B1, B2, B3, S, M, L. We optimised the smaller variants to be faster and the larger ones that are S,M,L at the expense of increased complexity. This experiment allowed us to make a suitable balance between accuracy and efficiency.

D. Training Process

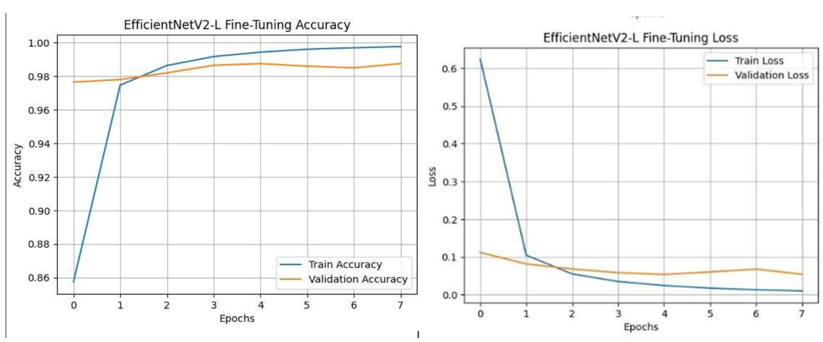

Figure –2-3 Training Accuracy and Training loss of EfficientNet V2-L variant

The training involves 2 stages: initial training and fine tuning. Firstly in the first stage, to preserve features learnt from imagenet dataset, the pretrained model was frozen. the classification head was changed to 15 classes. Adam optimizer was used, and sparse categorical cross entropy for multiclass classification. Also at the end the layer applied was SoftMax activation function [20] to the outputs. The batch size was taken as 32 for upto 10 epochs.
During the second phase, a significantly lower learning rate of 1e-5 was applied. The loss functions and evaluation metric used in 1st phase were used again along with the early stopping mechanism for monitoring

validation performance. This two-stage training process, that combined the stability of transfer learning with the flexibility of fine-tuning, helped the EfficientNetV2 models to achieve high performance and generalization for the task of predicting endangered species.

IV. RESULT

On the basis of training and validation loss, and accuracy, the performance of the different EfficienceNetV2 model variants (B0,B1,B2 ,B3,S,M,L) was evaluated. The results are given in Table-1.
The strongest performance was shown by EfficientNetV2-B0 with results of training accuracy as 97.28% and validation accuracy of 96.80%. The results suggested that as the model size increased, the accuracy of the model used also increased. EfficientNetV2-B1 yields training accuracy of 97.46% and a validation accuracy of 96.95% and EfficientNetV2-B2 further enhanced performance with a training accuracy of 97.76% and a validation accuracy of 97.35%.
As per the findings, as larger models were used, accuracy became better .As far as variant EfficientNetV2-B3 is concerned, it showed training accuracy of 97.88% and a validation accuracy of 96.95% which were slightly lower than the previous models.
EfficientNetV2-S, M, and L variants have higher accuracy as they are larger models. For endangered species prediction, the EfficientNetV2-S model yielded training accuracy of 99.10%. EfficientNetV2-M model shows improvement training accuracy to 99.49% and 97.95% validation accuracy. EfficientNetV2-L achieved training accuracy of 99.71%.
EfficientNetV2-B0 variant, which had training loss of 0.1002, had high accuracy, and it shows small models can also perform well. EfficientNetV2-L achieved the lowest loss of 0.0121.
A relation was found that larger models had high accuracy but the smaller models also had strong performance. EfficientNetV2-L model shows the best performance in terms of both training and validation accuracy.

TABLE II: RESULTS OF THE VARIANTS OF EFFICIENTNET V2

Model Variant	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
EfficientNetV2	0.4062	0.8744	2.3264	0.5620
B0	0.1002	0.9728	0.1129	0.9680
B1	0.0930	0.9746	0.1177	0.9695
B2	0.0839	0.9776	0.0865	0.9735
B3	0.0749	0.9788	0.1223	0.9695
S	0.0334	0.9910	0.0890	0.9775
M	0.0201	0.9949	0.0913	0.9795
L	0.0121	0.9971	0.0542	0.9875

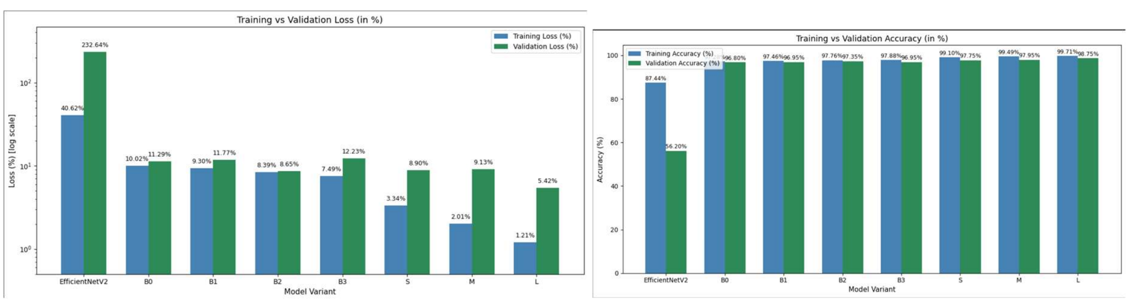

Figure 4-5 Comparison of training and validation loss and accuracy percentages for various EfficientNetV2 model variants.

V. CONCLUSION

Current paper presents applications of deep learning techniques by using EfficientNetV2 variants to classify animals for biodiversity conservation. Particularly, the EfficientNetV2-L model achieved the highest performance, with a training accuracy of 99.71% and a validation accuracy of 98.75%
In the paper transfer learning was used on a labelled dataset of endangered species.
The system did a great job at predicting which species are in danger, but there’s still room to grow. We need to test it on more animals from different places to see if it works everywhere. Also, imagine if we could feed it more information, like where animals live, changes in the environment, and how populations are doing. That would give us a much clearer picture and make it an even more powerful tool for protecting wildlife.

The main drawback of our paper is that only 5400 images across 90 animal species are used which do not represent all the animals. Future research can use larger models for improving this study.

To sum up, this research presents the potential of machine learning for advancing conservation efforts by accurately predicting endangered species and providing insights into biodiversity trends. As AI continues to evolve, integrating machine learning models into real-world conservation projects could significantly enhance our ability to protect endangered species and preserve global biodiversity.

REFERENCES

[1] R. Fernandes, C. Premebida, P. Peixoto, D. Wolf and U. Nunes, "Road Detection Using High Resolution LIDAR," 2014 IEEE Vehicle Power and Propulsion Conference (VPPC), Coimbra, Portugal, 2014, pp. 1-6, doi: 10.1109/VPPC.2014.7007125. keywords: {Roads;Laser radar;Cameras;Benchmark testing;Three-dimensional displays;Robots;Morphological operations},

[2] Jiménez López, J., & Mulero-Pázmány, M. (2019). Drones for conservation in protected areas: present and future. Drones, 3(1), 10.

[3] Badhe, T., Borde, J., Waghmare, B., Thakur, V., & Chaudhari, A. (2022). Study of Deep Learning Algorithms to Identify and Detect Endangered Species of Animal. International Research Journal of Engineering and Technology, 11.

[4] Modi, S. ARE THE CURRENT EFFORTS ENOUGH TO CONSERVE WILDLIFE?.

[5] Gray, P. C., Fleishman, A. B., Klein, D. J., McKown, M. W., Bezy, V. S., Lohmann, K. J., & Johnston, D. W. (2019). A convolutional neural network for detecting sea turtles in drone imagery. Methods in Ecology and Evolution, 10(3), 345-355.

[6] Rominger, K. R., & Meyer, S. E. (2021). Drones, deep learning, and endangered plants: A method for population-level census using image analysis. Drones, 5(4), 126.I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[7] Kellenberger, B., Marcos, D., & Tuia, D. (2018). Detecting mammals in UAV images: Best practices to address a substantially imbalanced dataset with deep learning. Remote sensing of environment, 216, 139-153. R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[8] Hua, A., Martin, K., Shen, Y., Chen, N., Mou, C., Sterk, M., ... & Jewell, Z. C. (2022). Protecting endangered megafauna through AI analysis of drone images in a low-connectivity setting: a case study from Namibia. PeerJ, 10, e13779.

[9] Wich, S. A., Hudson, M., Andrianandrasana, H., & Longmore, S. N. (2021). Drones for conservation. Conservation technology, 35.

[10] Zhang Z. Drone-YOLO: An Efficient Neural Network Method for Target Detection in Drone Images. Drones. 2023; 7(8):526. https://doi.org/10.3390/drones7080526

[11] Bondi, E., Kapoor, A., Dey, D., Piavis, J., Shah, S., Hannaford, R., ... & Tambe, M. (2018, July). Near Real-Time Detection of Poachers from Drones in AirSim. In IJCAI (pp. 5814-5816).

[12] Xu, Z., Shi, H., Li, N., Xiang, C., & Zhou, H. (2018, November). Vehicle detection under UAV based on optimal dense YOLO method. In 2018 5th international conference on systems and informatics (ICSAI) (pp. 407-411). IEEE.

[13] Chalmers, C., Fergus, P., Wich, S., & Montanez, A. C. (2019). Conservation AI: Live stream analysis for the detection of endangered species using convolutional neural networks and drone technology. arXiv preprint arXiv:1910.07360.

[14] Kaur, A., Kukreja, V., Chattopadhyay, S., Verma, A., & Sharma, R. (2024, March). Empowering Wildlife Conservation with a Fused CNN-SVM Deep Learning Model for Multi-Classification Using Drone-Based Imagery. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI) (Vol. 2, pp. 1-4). IEEE.

[15] Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946. https://arxiv.org/abs/1905.11946

[16] Inazu, Y., & Kimata, H. (2023, June). Study on Lightweight and Efficient Deep Image Compression Method Based on MBConv. In Proceedings of the 2023 7th International Conference on Graphics and Signal Processing (pp. 35-41).

[17] Koonce, B. (2021). EfficientNet. In Convolutional neural networks with swift for Tensorflow: image recognition and dataset categorization (pp. 109-123). Berkeley, CA: Apress.

[18] V. -T. Hoang and K. -H. Jo, "Practical Analysis on Architecture of EfficientNet," 2021 14th International Conference on Human System Interaction (HSI), Gdańsk, Poland, 2021, pp. 1-4, doi: 10.1109/HSI52170.2021.9538782. keywords: {Computer vision;Analytical models;Costs;Computational modeling;Computer architecture;Real-time systems;Mobile handsets;efficient CNN architecture;efficientnet;SE module},

[19] https://www.kaggle.com/datasets/iamsouravbanerjee/animal-image-dataset-90-different-animals

[20] A. Kagalkar and S. Raghuram, "CORDIC Based Implementation of the Softmax Activation Function," 2020 24th International Symposium on VLSI Design and Test (VDAT), Bhubaneswar, India, 2020, pp. 1-4, doi: 10.1109/VDAT50263.2020.9190498. keywords: {Timing;Table lookup;Field programmable gate arrays;Neuromorphics;Machine learning;Hardware;Biological neural networks;Neuromorphic Circuits;Softmax Activation;Deep Learning;Digital Design},