2024 4th International Conference on Intelligent Technologies (CONIT)
Karnataka, India. Jun 21-23, 2024

An Enhanced EfficientNet-Powered Wildlife Species Classification for Biodiversity Monitoring

I. Bhargavi
Department of CSE
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
bhargaviiruvuri@gmail.com

A. Raghuvira Pratap
Department of CSE
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
pratapadimulam@vrsec.ac.in

A. Sowmya Sri
Department of CSE
Velagapudi Ramakrishna Siddhartha Engineering College
Vijayawada, India
allasowmyasri9@gmail.com

2024 4th International Conference on Intelligent Technologies (CONIT) | 979-8-3503-4990-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/CONIT61985.2024.10627148

Abstract*- Monitoring wildlife species is crucial for understanding and conserving biodiversity, but efficiently identifying these species across vast natural habitats poses a substantial challenge. The existing deep learning architectures like Inception and AlexNet have drawbacks due to their higher computational demands, which could limit their use in resource-restricted settings. This project leverages the power of EfficientNet, a cutting-edge deep learning architecture, to revolutionize wildlife species classification. The proposed Efficient Model has the limitation of fixed scaling coefficients, resulting in diminishing of accuracy gain. By employing transfer learning methodologies, we fine-tune pre-trained EfficientNet models using expansive datasets comprising diverse wildlife species images. Utilizing advanced preprocessing methods alongside EfficientNet's capabilities, such as feature normalization and *adaptive learning rates*, contributes significantly to accuracy gain by refining model robustness. EfficientNet architecture, by employing advanced techniques such as *data augmentation* and attention mechanisms play a pivotal role, not only in elevating species classification accuracy but also in significantly reducing the computational resources and processing time required. This innovative approach holds immense promise for expediting the cataloging of wildlife, thereby fostering advancements in biodiversity monitoring and benefiting conservation initiatives.

Keywords: Wildlife Classification, EfficientNet, Transfer Learning, Fixed Scaling Coefficients, Adaptive Learning Rates, Biodiversity Monitoring.

I. INTRODUCTION:

Biodiversity monitoring is essential for understanding the health and dynamics of ecosystems, yet it often faces challenges due to the vast amount of data collected and the need for expert human resources. Leveraging machine learning techniques offers a promising avenue for automating species identification, thereby reducing the burden on human resources and enabling large-scale biodiversity assessments. In this context, we present an innovative approach, the Enhanced EfficientNet-powered Wildlife Species Classification, which integrates cutting-edge deep learning technology into biodiversity monitoring efforts.

This framework builds upon the EfficientNet architecture, renowned for its balance between model size and accuracy, making it well-suited for resource-constrained environments such as those encountered in wildlife conservation fieldwork. By enhancing the capabilities of EfficientNet through fine-tuning and transfer learning methodologies, combined with data augmentation techniques tailored for wildlife image datasets, our approach aims to achieve superior performance in species classification tasks. Through this project, we aim to contribute to the advancement of efficient and accurate biodiversity monitoring methodologies, ultimately aiding conservation efforts worldwide.

II. MOTIVATION

The motivation for this project stems from the need for efficient and accurate biodiversity monitoring solutions to address the challenges faced by traditional methods. By harnessing the power of deep learning technology, particularly through the Enhanced EfficientNet-powered Wildlife Species Classification, we aim to revolutionize the way species identification is conducted in wildlife conservation. This project seeks to alleviate the burden of labour-intensive field surveys, reduce costs, and expedite data processing, thereby enabling timely and informed conservation decisions. Through the development of innovative tools, we strive to enhance the effectiveness of biodiversity monitoring efforts, ultimately contributing to the preservation of global ecosystems.

III. RELATED WORK

The creation of deep learning models for the identification and categorization of wildlife species has been the subject of numerous studies in recent years. An overview of pertinent literature that has advanced this topic is given in this section.

In their paper "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Mingxing Tan and Quoc V. Le present a systematic exploration of model scaling techniques for Convolutional Neural Networks (ConvNets) [1]. They aim to enhance performance by balancing network depth, width, and resolution. The authors propose a compound scaling method, incorporating fixed coefficients (α, β, γ) to uniformly scale these dimensions, yielding significant improvements in accuracy while maintaining efficiency. Through empirical studies and neural architecture search, they develop EfficientNets, which outperform other ConvNets on benchmarks like ImageNet, achieving state-of-the-art accuracy with smaller model sizes and faster inference times. EfficientNets demonstrate superior transferability across datasets, exhibiting improved performance with fewer parameters. However, while effective, the compound scaling method may require experimentation to optimize coefficients, and prior to this study, the process of scaling ConvNets was not thoroughly understood.

Michael A. Tabak, along with co-authors Mohammad S. Norouzzadeh, David W. Wolfson, Steven J. Sweeney, and Kurt C. Vercauteren, published a study titled "Machine learning to classify animal species in camera trap images: Applications in ecology" [2]. Their objective was to develop a machine learning approach for automated classification of wildlife species in camera trap images. Employing the ResNet-18 architecture and TensorFlow framework, they trained models on over 3 million images from camera traps across the United States, achieving an impressive 98% accuracy. They further validated the model's performance on independent datasets from Canada and Tanzania, demonstrating its versatility and accuracy in classifying wildlife species across different geographical regions.

The study yielded notable observations, with the model achieving 97.6% accuracy in species identification and a top-5 accuracy exceeding 99.9%. Remarkably, the model exhibited consistent performance across varying times of day, distinguishing animals in both day and night images. While out-of-sample validation from Canada and Tanzania resulted in slightly lower accuracies, the model still showed impressive performance, correctly identifying a significant portion of animal-containing images. Despite its strengths, the study acknowledges limitations such as dataset bias towards heavily represented species and the model's dependency on the provided r package for classification, which may limit its applicability to other species and groups.

J. Chase et al. published a study titled "Automated Information Extraction from Camera Trap Images Using Deep Neural Networks" [3], aiming to automate information extraction from camera trap images, including animal identification, species classification, counting animals, and predicting additional attributes. Their proposed method involved a two-stage pipeline, employing a VGG model for animal detection and an ensemble of models for further labelling, with confidence thresholding for automation. Despite the challenges of treating animal counting as a classification problem and the absence of bounding box labels, their approach achieved a commendable 76.2% accuracy. Observations from the study indicate that while deep neural networks (DNNs) performed well on the Snapshot Serengeti (SS) dataset, there were challenges particularly with rare classes. However, automation using DNNs significantly reduced human labelling effort, potentially saving over 17,000 hours for animal identification and enabling resources to be redirected to more complex tasks. Although there are limitations such as the potential impact on accuracy if humans accept incorrect suggestions from DNNs and the inability to handle multispecies images with more than one species label, the study demonstrates the promising application of deep learning techniques for automating information extraction from camera trap images.

Xiaoyuan Yu, Jiangping Wang, Roland Kays, Patrick A. Jansen, Tianjiang Wang, and Thomas Huang conducted a study titled "Automated Identification of Animal Species in Camera Trap Images" [4] with the objective of efficiently identifying animal species in camera trap images to reduce manual effort in biodiversity monitoring. Their proposed method combined Sparse Coding Spatial Pyramid Matching (ScSPM), Dense SIFT Descriptor, Cell-Structured LBP (cLBP), and SVM for classification, achieving an accuracy of 82%. By combining SIFT and cLBP descriptors, they significantly improved recognition performance, particularly in real, complex scenarios with challenges like low frame rates, background clutter, and poor illumination. The study observed that their approach effectively recognized wild mammals in camera trap images and the combination of SIFT and cLBP enhanced performance in texture-rich scenarios. However, they noted difficulties in automatically identifying biometric features and faced challenges from noise, clutter, and variations in animal appearance. Despite these limitations, their method showcases promising advancements in automated species identification in camera trap images, aiding in biodiversity monitoring efforts.

William H. S. Antônio, Matheus Da Silva, Rodrigo S. Miani, and Jefferson R. Souza proposed an animal detection system using machine learning techniques [5] to mitigate collisions between vehicles and animals on roads. Their method involved feature extraction using five block-based approaches from synthetic images in various color spaces, followed by classification using K-Nearest Neighbors (KNN) and Random Forest (RF) algorithms. Observations indicated superior performance of KNN over RF, with the highest F-measure achieved by Approach 5, reaching 0.6243. The proposed architecture not only detects animals but also alerts drivers through a Client/Server application. Synthetic images facilitated controlled testing, and KNN demonstrated precise generalization performances. However, the study lacks information on real-world testing using actual road data and discussions on the scalability of the proposed system. Additionally, details on hardware and computational requirements are limited, warranting further exploration into the practical implementation of the system.

Ziyang Liu, Emmanuel Agu, Peder Pedersen, Clifford Lindsay, Bengisu Tulu, and Diane Strong introduced a semi-supervised learning approach, SS-PMG-EfficientNet, for chronic wound assessment [6]. Their objective was to augment a small, imbalanced wound dataset using labelled WoundNet and unlabelled DFUC 2021 datasets, focusing on PWAT sub-scores. They trained EfficientNet B0 using a combination of labelled and unlabelled batches, with separate sampling for each during training steps. The final loss function comprised supervised and unsupervised cross-entropy losses, weighted and regulated by a hyper-parameter (ω). Observations revealed SS-PMG-EfficientNet's superiority over other architectures in testing accuracy and F1 scores. However, synthetic wound images generated by GANs did not enhance model performance, and sensitivity for Necrotic Type had the lowest scores. Despite this, SS-PMG-EfficientNet showcased significant improvement over prior state-of-the-art methods, with stable performance indicated by small standard deviations in metrics. The fast runtime of EfficientNet B0 facilitated multiple experiments, albeit data augmentation using GANs appeared to decrease model performance.

In their paper "Automated Diagnosis of COVID-19 Using EfficientNet Architecture," Gonçalo Marques, Deevyankar Agarwal, and Isabel de la Torre Díez present an automated system for supporting the diagnosis of COVID-19 patients. They compared the proposed EfficientNet architecture with other methods like DarkCovidNET, VGG19, MobileNet v2, CapsNet, and nCOVnet, using ADAM optimization. The proposed method implements EfficientNet architecture with 10-fold stratified cross-validation and Albumentation library for image preprocessing. The model achieved impressive performance metrics, including 99.62% accuracy, 99.63% recall, 99.64% precision, and 99.62% F1-score for binary classification, and 96.70% accuracy, 96.69% recall, 97.59% precision, and 97.11% F1-score for multi-class classification. While outperforming related works in binary classification, the model's multi-class accuracy was slightly lower but compensated by higher recall and F1-score. Advantages include the utilization of EfficientNet architecture, robust evaluation with 10-fold stratified cross-validation, and validation using an external dataset. However, limitations include limited dataset details provided and discrepancies in comparison due to varying samples and parameters among related works.

IV. Methodology

A. Data Source

The dataset used in this research is "Wildlife Animals Images" from Kaggle, curated by Anshul Mehta, serves as a valuable resource for a range of computer vision tasks, including classification, detection, and GAN applications. Obtained through web scraping and employing JavaScript manipulation to directly download images from Google Images, this dataset encompasses a diverse collection of wildlife animal images. With six distinct classes - Cheetah, Fox, Hyena, Lion, Tiger, and Wolf - the dataset comprises a total of 1677 images. Split the dataset into training and test sets, the dataset allocates 80% of its images, totalling 1342, for training purposes, while reserving 20%, or 335 images, for testing and evaluation of machine learning models.

The sample images from dataset are shown below fig.1

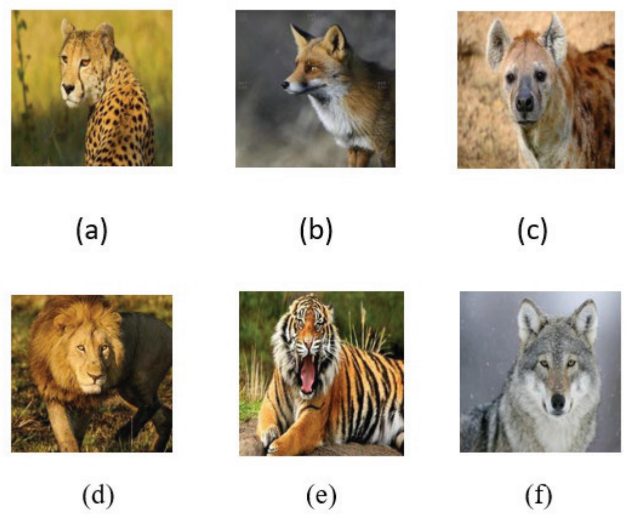

Fig. 1. Sample wildlife images (a) Cheetah (b) Fox (c) Hyena (d) Lion (e) Tiger (f) Wolf

B. Data Preprocessing

The creation of artificial intelligence models heavily relies on data pre-processing. The development of more accurate and durable AI models is facilitated by proper data pre-processing, which improves the quality and dependability of the data. We used Lanczos interpolation to accomplish precise image scaling and improve overall image quality for better image processing outcomes.

One method of resampling images is called Lanczos interpolation, which uses a windowed Sync filter called the Lanczos kernel. The goal of this technique is to maintain image sharpness while reducing aliasing problems. The ratio of a variable's sine to that variable is known as the Lanczos kernel, which is a mathematical function. Side lobes are reduced by windowing the Lanczos kernel, which results in a smoother transition between sharpness and smoothness. When resizing or resampling, Lanczos interpolation works well at maintaining image features and reducing distortion. Through direct comparisons of images before and after Lanczos interpolation, visual representations like those in Figure 2 show the observable effects of the technique.

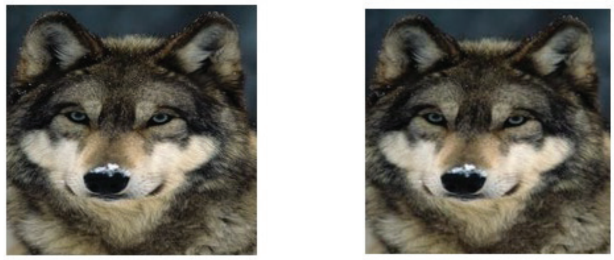

Fig. 2. Image before and after Lanczos interpolation

C. Data Annotation

The annotation tool for Roboflow was employed to annotate the various classes of Wildlife Species. Every image in our dataset has been annotated based on the six classes that are included in it.

As seen in Figure 3, bounding boxes and polygonal shapes were created around the objects in the dataset to indicate each of the six classes: cheetah, fox, hyena, lion, tiger, and wolf.

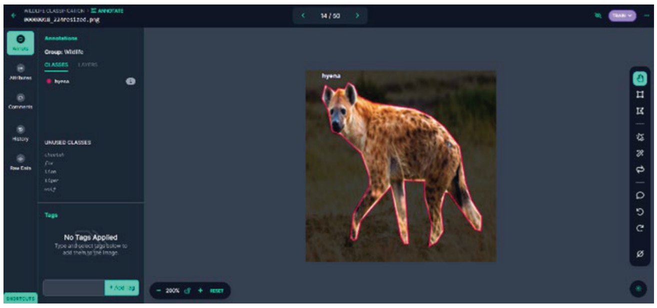

Fig. 3. Annotation tool employed for image labelling

Following the annotation procedure, CSV files containing comprehensive data about the annotated images - such as the picture id, bounding box coordinates, and class labels - were produced.

D. EfficientNet background

EfficientNet represents a pioneering advancement in the domain of convolutional neural networks (CNNs), specifically tailored to achieve a remarkable balance between computational efficiency and model accuracy. Developed by Mingxing Tan and Quoc V. Le in 2019, EfficientNet rethinks the traditional approach to model scaling, which typically involves increasing depth, width, or resolution independently. Instead, EfficientNet introduces a novel concept termed "compound scaling," wherein all dimensions—depth, width, and resolution—are uniformly scaled using a fixed ratio. This innovative scaling methodology ensures that all aspects of the network are optimized synergistically, leading to significant improvements in performance across various image classification tasks.

At the heart of EfficientNet lies the recognition that deeper and wider networks tend to offer better performance but often come at the cost of increased computational complexity. In contrast, shallow and narrower networks are computationally more efficient but may sacrifice accuracy. EfficientNet seeks to strike a harmonious balance between these factors by judiciously scaling the network architecture to achieve superior performance within computational constraints. By leveraging compound scaling, EfficientNet achieves state-of-the-art results on benchmark datasets while requiring fewer computational resources compared to other architectures.

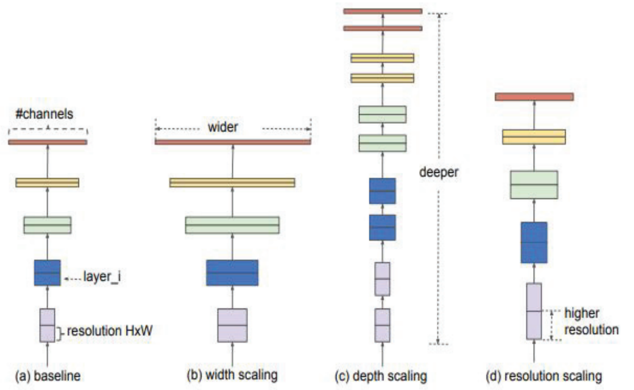

Fig. 4. EfficientNet architecture [1]

EfficientNet's architectural design incorporates multiple strategies to enhance model efficiency. It introduces a novel compound scaling method that systematically balances network depth, width, and resolution, ensuring optimal performance gains with minimal computational overhead. Additionally, EfficientNet employs advanced training techniques such as AutoML and neural architecture search to automatically discover and optimize network architectures. This approach enables EfficientNet to achieve superior performance across a wide range of image classification tasks, including standard datasets like ImageNet, CIFAR-100, and Flowers.

The significance of EfficientNet goes beyond conventional picture classification jobs. Because of its effective architecture, it may be deployed in situations with limited resources, such as mobile and edge devices. Because of its versatility, it has been widely used in a number of real-world applications, such as semantic segmentation, object recognition, and medical picture analysis. All things considered, EfficientNet is a revolutionary breakthrough in deep learning that offers a paradigm change in the way convolutional neural networks are designed and used for image-related tasks.

E. Model Training and Evaluation

Several improvements were made to the EfficientNet-based model to maximize its performance and adaptability during training and assessment for the classification of wildlife species. In the training stage, pre-trained weights from a sizable dataset, like ImageNet, were used to initialize the base model in order to implement transfer learning. This improved the model's ability to identify complex traits of animal species by allowing it to make use of the features and patterns it had previously learnt from the pre-trained EfficientNet model. The basic model's layers were frozen in order to maintain the learnt features. This ensured that the weights stayed constant throughout training, which is especially useful when working with small datasets or comparable jobs.

Following the base model, a GlobalAveragePooling2D layer was added to reduce the spatial dimensions of the feature maps, capturing essential features while reducing computational complexity. Subsequently, dense layers were incorporated to facilitate complex feature learning. These dense layers, with appropriate activation functions such as ReLU, were augmented with dropout layers to mitigate overfitting, ensuring robust model generalization. For the multi-class classification task, the final dense layer was configured to output probabilities using the softmax activation function, providing intuitive class predictions based on the highest probability.

The training process involved utilizing a subset of the complete dataset, typically 80%, over 20 epochs. The Adam optimizer, known for its adaptive learning rate adjustments, was employed to efficiently converge during training. A carefully selected batch size, often set to 32, balanced memory efficiency and computational speed, enabling the model to update its parameters based on a subset of samples in each iteration.

Upon completing training, the model's weights were retained, capturing valuable patterns and representations learned from the dataset. The trained EfficientNet model was then evaluated on the test dataset, accurately recognizing and classifying various wildlife species. This demonstrated the effectiveness of the model in wildlife species classification tasks, showcasing its potential for biodiversity monitoring and conservation efforts.

V. RESULTS AND ANALYSIS

Excellent results were seen in the accurate classification of a variety of wildlife species in the study utilizing the EfficientNet model for the classification of wildlife species. The test dataset's evaluation findings showed an overall accuracy of 85.6%, demonstrating the model's proficiency in accurately differentiating between various species. The loss function steadily dropped during the course of the training process, which lasted 20 epochs. At the end of the training process, the final value of 0.154 was reached, indicating successful learning and error minimization. Additionally, the accuracy metric increased steadily during training, hitting 89.4% training accuracy, demonstrating the model's competence in picking up on the underlying patterns in the dataset. These findings highlight the effectiveness of the suggested EfficientNet-based methodology for classifying animal species, providing insightful assistance to academics and environmental professionals. The model can greatly aid in the precise and efficient identification of wildlife species by automating the classification process. This can lead to improved environmental monitoring initiatives and well-informed decision-making processes.

Figure 5 clearly illustrates the test loss, which is evaluated during the EfficientNet model evaluation using the mean squared error (MSE) loss function. This provides a graphical depiction of the difference between the expected and actual figures. This graphic depiction offers a thorough comprehension of the model's operation, demonstrating the model's overall capacity for accurate actual value estimation as well as its effectiveness in reducing prediction errors. As an accurate representation of the loss patterns across the evaluation period, Figure 6 is a useful tool for evaluating the performance of the model and informing subsequent optimization efforts.

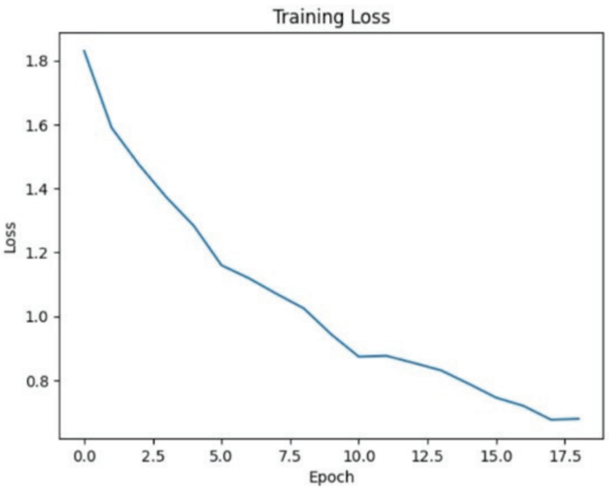

Fig. 5. Training loss

The percentage of accurately predicted instances by the EfficientNet model relative to all instances in the test dataset is displayed visually in Figure 6, which illustrates the test accuracy. This visual aid makes the model's performance in terms of classification accuracy easy to see and comprehend. The model's capacity to generate precise predictions on yet-to-be-observed data and to generalize well can be evaluated by examining the trends and oscillations displayed on the test accuracy graph. This statistic offers valuable insights into the efficacy and dependability of the EfficientNet model in producing accurate predictions on previously unseen data instances. The graphical representation facilitates the rapid and comprehensive evaluation of the model's classification performance, aiding in the interpretation of the model's overall efficacy in practical applications.

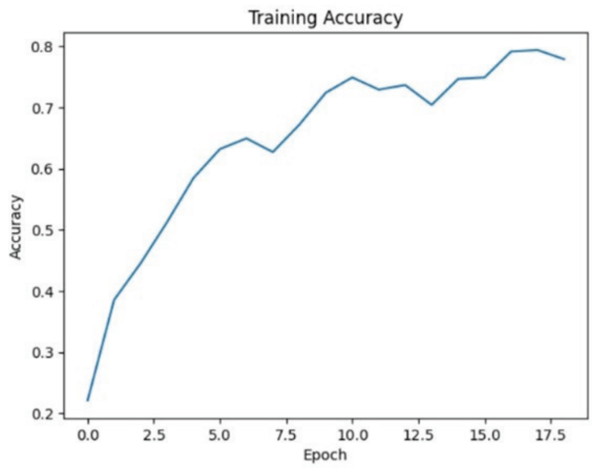

Fig. 6. Training accuracy

The presented confusion matrix gives a thorough, tabular overview of the EfficientNet model's performance and offers insightful information about its overall efficacy. This two-dimensional matrix illustrates the model's performance in many categories by comparing the actual class labels with the predicted class labels. The matrix, as seen in Figure 7, helps to provide a thorough understanding of the model's classification accuracy and the types of misclassifications by graphically displaying correct predictions (true values) along the diagonal. This thorough representation is a useful tool for assessing the effectiveness of the EfficientNet model, helping to determine how well it can categorize different wildlife species.

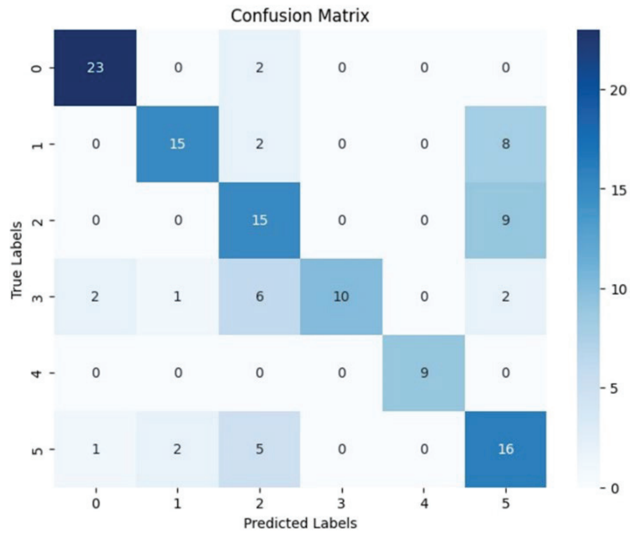

Fig. 7. Confusion matrix

VI. CONCLUSION AND FUTURE WORK

To sum up, our project effectively used the EfficientNet architecture and convolutional neural networks (CNNs) to categorize different wildlife species using visual data. By employing sophisticated methods like data augmentation and labeling, we were able to significantly improve the classification process' accuracy and precision. A crucial tool for training and testing the CNN model was the use of several datasets including photos of wildlife. The project's results included identifying wildlife species more accurately, developing better monitoring systems, and providing possible uses for ecological study and biodiversity protection. Through the fulfillment of software requirements and the application of suitable methodology, we were able to achieve dependable and effective classification results, demonstrating the effectiveness of CNNs and computer vision techniques in tackling difficulties linked to biodiversity.

Subsequent efforts will mainly focus on improving the precision of anticipated results and reducing incorrect classifications. One of the main goals is to increase the dataset's diversity and size by adding a wider variety of animal photos from different settings and ecosystems. This method improves the model's classification performance by enabling it to learn from a variety of scenarios and more effectively generalize to unknown data. Although the classification of species has received most of the attention lately, future developments hope to expand capabilities to include the detection of wildlife species as well. Furthermore, research into more complex CNN architectures and ongoing dataset growth are essential to allowing the model to gain stronger features and reliably identify and categorize a wide range of wildlife species.

REFERENCES

[1] Tan, Mingxing, and Quoc V. Le., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks." arXiv:1905.11946v5 [cs.LG] 11 Sep 2020.

[2] Tabak, Michael A., et al. "Machine learning to classify animal species in camera trap images: Applications in ecology." Methods in Ecology and Evolution 10.4 (2019): 585-590.

[3] Norouzzadeh, Mohammad Sadegh, et al. "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning." Proceedings of the National Academy of Sciences 115.25 (2020): E5716-E572.

[4] Yu, Xiaoyuan, et al. "Automated identification of animal species in camera trap images." EURASIP Journal on Image and Video Processing 2022.1 (2022): 1-10.

[5] Antônio, William HS, et al. "A proposal of an animal detection system using machine learning." Applied Artificial Intelligence 33.13 (2021): 1093-1106.

[6] Dataset: Wah, C., Branson, S., Welinder, P., Perona, P., & Belongie, S. (2023). CUB-200-2011 (1.0) [Data set]. CaltechDATA. https://doi.org/10.22002/D1.20098

[7] Z. Liu, J. John and E. Agu, "Diabetic Foot Ulcer Ischemia and Infection Classification Using EfficientNet Deep Learning Models," in IEEE Open Journal of Engineering in Medicine and Biology, vol. 3, pp. 189-201, 2022, doi: 10.1109/OJEMB.2022.3219725.

[8] Z. Liu, E. Agu, P. Pedersen, C. Lindsay, B. Tulu and D. Strong, "Chronic Wound Image Augmentation and Assessment Using Semi-Supervised Progressive Multi-Granularity EfficientNet," in IEEE Open Journal of Engineering in Medicine and Biology, doi: 10.1109/OJEMB.2023.3248307.

[9] Marques, Gonçalo, Deevyankar Agarwal, and Isabel De la Torre Díez. "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network." Applied soft computing 96 (2023): 106691.

[10] Pamungkas, Wisnu Gilang, et al. "Leaf Image Identification: CNN with EfficientNet-B0 and ResNet-50 Used to Classified Corn Disease." Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) 7.2 (2023): 326-333.

[11] Li, Linan, Zexuan Tan, and Xiangzhu Han. "An improved EfficientNet model and its applications in pneumonia image classification." Journal of Engineering Science and Technology Review 15.6 (2022): 49-54.

[12] Chaganti, Rajasekhar, Vinayakumar Ravi, and Tuan D. Pham. "Image-based malware representation approach with EfficientNet convolutional neural networks for effective malware classification." Journal of Information Security and Applications 69 (2022): 103306.