Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 11
November 2024

Improved Convolutional Neural Networks Model in the Identification and Classification of Monkey Species

***Benjamin Asubam Weyori[1], Emmanuel Harris[2], Bright Karim-Abdallah[3], Opoku Gyabaah[4], Gyimah Kopri[5, 6]**

[1]Department of Computer and Electrical Engineering, University of Energy and Natural Resources, Sunyani, Ghana, * benjamin.weyori@uenr.edu.gh

[2]Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi – Ghana

[3]Quality Assurance and Academic Planning Directorate, University of Energy and Natural Resources, Sunyani, Ghana

[4]Department of Mathematics and Statistics, Ghana Communication Technology University, Accra – Ghana

[5]Department of Electrical Engineering, Sunyani Technical University, Sunyani, Ghana

[6]Department of Computer Science and Informatics, University of Energy and Natural Resources Sunyani, Ghana

Abstract

This study demonstrates the utilisation of the Inception V3 model, a deep Convolutional Neural Network (CNN), for automatically identifying and categorising monkey species through computer vision. Our research seeks to improve monitoring and classification by applying modern machine-learning methodologies in light of the essential need for precise species identification for wildlife conservation. The dataset, consisting of images from 10 distinct monkey species, was enhanced by mathematical transformations, including rotations, flips, and colour jittering to increase model generalisation. The Inception V3 model was trained for 50 epochs, attaining a maximum training accuracy of 100% and a peak validation accuracy of 93.75%. The validation loss showed variability, signifying overfitting and difficulties in generalising to unseen data. A comparative study with other research, including those by Brust et al. (2017) and Freytag et al. (2016), validates our model's competitive strength while reducing prevalent drawbacks, such as reliance on high-quality datasets and susceptibility to environmental fluctuations. To mitigate these constraints, we recommend for future works into techniques like multi-task learning (MTL), attention processes, and synthetic data generation employing Generative Adversarial Networks (GANs). Our results illustrate the capability of the Inception V3 model for real-time, automated wildlife surveillance, establishing a basis for more effective and scalable conservation strategies. This research enhances the precision and efficacy of species identification in dynamic and harsh habitats, hence aiding in the conservation of endangered species.

Keywords: Monkey Species Classification, Inception V3, Wildlife Conservation, Deep Learning, Ecology

1.0 Introduction

Biodiversity monitoring, ecological research, and conservation initiatives rely on precise animal species identification and categorisation (Shukla et al., 1878). Because of their ecological importance and the serious dangers they confront, primates are especially important subjects for this type of identification. Finding effective, non-invasive ways to track primate populations is critical because their numbers are falling due to habitat degradation, poaching, and global warming(Rajalingham et al., 2018). When dealing with big datasets or morphologically similar species, the labour-intensive and expensive traditional methods of visual examination and genetic analysis are prone to errors. In particular, deep learning methods and convolutional neural networks (CNNs) have recently attracted much attention as potential solutions to these problems. CNNs have shown promising results in several animal classification tasks, thanks to their ability to extract powerful features (Brust et al., 2020; (Pawara et al., 2020).The performance of CNNs in animal identification has been demonstrated in numerous studies. For instance, a model that implemented the VGGNet architecture could accurately identify individual giant pandas with 95% precision, regardless of the quality of the image (Guan et al., 2023) .

Similarly, a CNN-based model was implemented in a study on the classification of monkey species, resulting in an accuracy of 99.44% in identifying individual primates from a dataset of 16,226 images. This success demonstrates the potential of machine learning to mitigate human error in ecological research (Shi et al., 2018). Nevertheless, the classification of similar-looking species, such as snub-nosed monkeys, continues to be challenging due to the subtle morphological differences that conventional CNN models cannot capture (Zhao et al., 2021). These constraints have been addressed by investigating enhancements such as multi-view CNNs for audio signals (Xu et al., 2020), modular architectures for optimising image recognition tasks (Recognition, 2022), and the utilisation of attention mechanisms to concentrate on discriminative regions of an image (Brust et al., 2020). However, challenges in generalisation and robustness continue to plague these methods. Various strategies have been developed by the research community to address these challenges, such as One-vs-One (OvO) classification schemes to simplify decision boundaries in multi-class problems (Pawara et al., 2020) and transfer learning, which involves fine-tuning pre-trained models like ResNet-50 and Inception-V3 on specific datasets (Trapanotto et al., 2022). Despite these developments, numerous extant models continue to be reliant on high-quality datasets and are unable to generalise effectively to unseen data, particularly when images are captured in varying environmental conditions with a low resolution (Rajalingham et al., 2018).

Additionally, models trained on particular datasets frequently fail to perform adequately on other datasets due to the complexities of field data, which frequently include occlusions noise and differences in data distribution (Schofield et al., 2019). To circumvent these constraints, this research aims to propose a novel methodology incorporating sophisticated techniques, i.e., transfer learning. This approach is anticipated to offer a more generalised and precise solution for identifying and classifying primate species, thereby providing valuable insights for conservation efforts. The proposed approach is designed to contribute substantially to the ongoing projects of wildlife conservation, particularly in deploying automated monitoring systems adaptable to real-world conditions, by addressing the challenges of overfitting, dataset imbalance, and the need for improved generalisation (Amran et al., 2024).

This study focuses on advancing ecological research and offers practical solutions to intricate wildlife monitoring duties by utilising advanced deep learning methods.

2.0 Problem formulation

In the field of biodiversity monitoring, the monitoring and classification algorithm requires critical and accurate performance. This aids in capturing and understanding the dynamic features of the ecological reserve. The problem statement governing this task is the similar-looking species of the large dataset(Shi et al., 2018). The proposed method's objective function is expected to follow the mathematical model shown.

Let the entire dataset be represented by,

$$\mathbb{D} = \{(x_i, y_i) \mid i = 1, \dots, N\} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (1)$$

Where $x_i \in R^{m\times n\times c}$ denotes the input image of an ith monkey species. Since the images are RGB coded, the dimension of each image is represented as $R^{m\times n\times c}$. With 10 monkey species, the label $y_i \in \{1, 2, \dots, 10\}$.

For a deep learning model $\mathcal{F}(\cdot\,;\,\Theta)$ is parametized by

$$\Theta = \{W, b\} \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots (2)$$

with W and b representing weight and bias terms across all layers of the model.

The objective is to search for the optimal parameter Θ_{opt}that minimizes the loss function $\mathcal{L}(\Theta)$. Here, the loss function measures the descrpancies between predicted outputs to reduce overfiting issues and aim at high level of generalizability.

Given the number of layers as L including pooling and fully connected layers, the feature extraction at each layer $l \in \{1, 2, \ldots, L\}$ is represented as:

$$h^{(l)} = f^{(l)}\left(W^{(l)} \times h^{(l-1)} + b^{(l)}\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Where $W^{(l)}$ and $b^{(l)}$ denote the weight and bias of layer l. $f^{(l)}(\cdot)$ represents the activation function applied at layer l.

The final output value of the fully connected layer applies a $softmax$ function that transforms the raw scores of the output into a probability distribution for each labelled class.

$$\hat{y}_i = softmax(z) = \frac{exp(z_j)}{\sum_{j=1}^{K} \exp(z_j)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots. (4)$$

Where $z = W^{(L)} \times h^{(L)} + b^{(L)}$

The loss function is hereby given by:

$$\mathcal{L}(\Theta) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{K=1}^{K} \hat{y}_{ik}\log(\hat{y}_{ik}) + \lambda R(\Theta) \ldots\ldots\ldots\ldots\ldots. (5)$$

Where $\hat{y}_{ik}$ denotes the binary indicator $(0, 1)$ when class K is classified correctly for the ith image.

$$R(\Theta) = \frac{1}{2}\sum_{l=1}^{L}\left(\left\|W^{(l)}\right\|_F^2 + \left\|b^{(l)}\right\|_2^2\right) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (6)$$

Equation (6) is introduced for regularization of the model. $\|\cdot\|_F^2$ and $\|\cdot\|_2^2$ are the Frobenium norm and L_2 norm respectively. In order to control the trade-off between data fitting and regularization, the value of λ is expected to be greater than 0.

3.0 Methodology

Implementing the Inception V3 model (Lin et al., 2019), a potent Convolutional Neural Network (CNN) pre-trained on the extensive ImageNet dataset (Krizhevsky & Hinton, 2012), and we present a transfer learning strategy to enhance the generalisability and accuracy of monkey species recognition from images. Because of its effective architecture and excellent feature extraction (Szegedy et al., 2016) capabilities, Inception V3 is a popular choice for handling challenging picture categorisation jobs. Through the process of fine-tuning the Inception V3 model on our particular dataset, which consists of 272 testing images and 1,102 training images from 10 different species of monkeys, we can efficiently handle issues with overfitting, limited data, and varying ambient conditions and image quality. Figure 3.1 shows a general block diagram of the Inception V3 architecture.

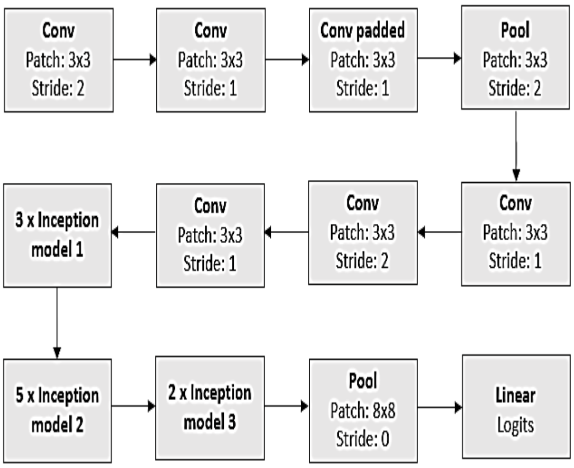


Figure 3.1: Architecture of a generic inception V3 model

Regarding the inception of the V3 model, the proposed approach entails creating new layers specifically for our classification assignment and replacing the last few layers of the model. The robust feature representations acquired from ImageNet will be preserved (Wang et al., 2019) since we can train only the newly created fully connected layers unique to classifying monkey species at first because the convolutional basis of Inception V3 will be frozen. The final few layers of the model will then be unfrozen and adjusted using a reduced learning rate to better match their features to the unique characteristics of our dataset. Data augmentation methods, including flipping, rotation, and colour jittering, will be used to diversify the training set of data to improve model performance even further. This method is anticipated to produce robust results with high classification accuracy in various environmental situations, making it possible to implement an automated, real-time wildlife monitoring system that will better assist conservation efforts.

3.1 Dataset Description

The 1,102 training and 272 testing images in the dataset, which represent ten distinct monkey species, were taken from the Kaggle Data Science community(Aldi et al., 2023).

Table 3.1: Names of all 10 species of monkeys in the dataset

Scientific Name	Common Name
Alouatta palliata	Mantled Howler Monkey
Aotus nigriceps	Black-headed Night Monkey
Cacajao calvus	Bald Uakari
Cebuella pygmaea	Pygmy Marmoset
Cebus capucinus	White-headed Capuchin
Erythrocebus patas	Patas Monkey
Macaca fuscata	Japanese Macaque
Mico argentatus	Silvery Marmoset
Rachypithecus johnnii	Nilgiri Langur
Saimiri sciureus	Common Squirrel Monkey

To avoid bias in model training, the dataset has a balanced distribution of species and each image is an RGB image with a range of resolutions. To comply with the input specifications of the Inception V3 model, images are uniformly sized to 224×224×224 by 224×224×224 pixels (Krizhevsky et al., 2012). Preprocessing techniques like normalisation and scaling were also applied to guarantee uniformity throughout the dataset. Figure 3.2 shows 4 samples of the images in the dataset.

the dataset

3.2 Data Preprocessing

In the data pre-processing stage, model generalisation enhancement was a key consideration. Data augmentation technology was hence applied to the training dataset. By applying the mathematical transformations with $\Theta \in [-20^0, 20^0]$ on each image, both vertical and horizontal flipping were achieved. Zooming and colour jittering (Zini et al., 2023) were also applied accordingly. The results of transformation were to diversify the training images. This will help expose the model to different categories of distortions and transformations of the individual images. One requirement of the inception V3 is to match the input scale with the images; hence, a normalisation scale of $[0, 1]$ was finally applied to the dataset.

3.3 Model training, fine-tuning and evaluation

The first phase of the transfer learning technique was to freeze the weights of the convolutional base of the inception V3 model. The frozen weights are represented as $\mathcal{F}(x;\ \Theta_{base})$. Here, only the newly connected layer $g(x;\ \Theta_{new})$ are trained on the dataset (Yosinski et al., 2014). The multiple output label calls for a categorical cross-entropy loss function $\mathcal{L}$, (Zhang, 2018) to optimize Θ_{new}. The optimization function is given by

$$\min(\Theta_{new})\ \mathcal{L}(y, H\ (x;\ \Theta_{base}),\ \Theta_{new}))$$

Where $H\ (x;\ \Theta_{base}),\ \Theta_{new})) = g(\mathcal{F}\ (x;\ \Theta_{base}),\ \Theta_{new}))$

Finally, we unfreeze the last 20 layers of the base model and fine-tune them with a learning rate lower than the initial learning rate. Here, 0.00001 was used further to adapt the pre-trained features to the present dataset. The training is conducted on Google Colab, utilising its complimentary GPU resources (including NVIDIA Tesla K80 GPUs) to expedite the computationally demanding procedure. Accuracy, precision, recall, and the F1-score, and the area under the receiver operating characteristic curve (AUC-ROC) are used to measure how well a model works (Powers, 2011). These measures give a full picture of how well the model can correctly classify the monkey species. A confusion matrix is also made to look at deviations in classifying different species. This lets us find any biases in the predictions.

4.0 Results and Discussion

The simulation shows a result of 50 epochs of training the Inception V3 model to identify and classify monkey species. We minimised the loss function for the validation and training datasets during training to maximise the model's accuracy. By the end of the training session, the model had achieved a flawless training accuracy of 100%, demonstrating substantial advances in accuracy. The difficulties in adapting the model to novel, previously unseen data were evident because validation accuracy fluctuated over epochs.

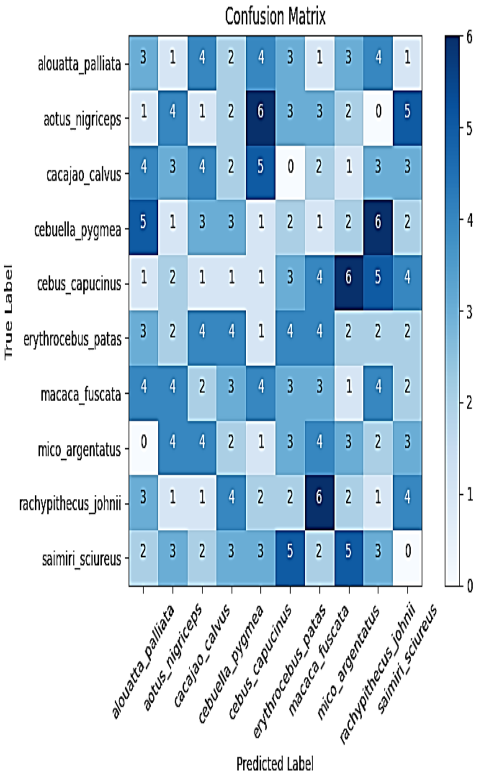
As shown in Figure 4.1, Training accuracy was 75.45%, and validation accuracy was 79.30% in the first epoch, with a loss of 0.7823 in training and a significant loss of 2.9472 in validation. These indicators show that the model had trouble applying what it learnt from the training data to the validation set. The model's accuracy in training increased dramatically during the training process, eventually reaching 100% by the second epoch and staying there for most of the future epochs. On the other hand, the validation accuracy varied from 62.5% to 93.75% during the epochs, indicating that the model experienced overfitting, a phenomenon in which it performed adequately on the training data but failed to generalise to the validation data. During epoch 34, the validation accuracy peaked at 93.75%, and the validation loss was 1.6405, suggesting that the model performed exceptionally well on the validation dataset.

(a)

(b)

Figure 4.1: visualization of training and validation accuracies (a) and losses (b)

The confusion matrix shown in figure 4.2 how well the model performed in categorising ten distinct species of monkeys. For each species, the diagonal elements show the correctly predicted instances, while the off-diagonal elements show incorrect classifications. The matrix indicates that the model has a great deal of trouble differentiating between a number of species, and numerous non-zero off-diagonal values imply that classification errors occur frequently. One common mix-up is *"cebuella_pygmea"* with *"cebus_capucinus"* and *"mico_argentatus,"* whereas *"saimiri_sciureus"* is incorrectly assigned to several different species. The species with the best accuracy, *"rachypithecus_johnii,"* has six right classifications. Other species, including *"aotus_nigriceps"*, also have some decent correct predictions but many misclassifications. The model shows significant potential for development, especially in terms of lowering cross-species confusion and raising classification accuracy.



	alouatta_palliata	aotus_nigriceps	cacajao_calvus	cebuella_pygmea	cebus_capucinus	erythrocebus_patas	macaca_fuscata	mico_argentatus	rachypithecus_johnii	saimiri_sciureus
alouatta_palliata	3	1	4	2	4	3	1	3	4	1
aotus_nigriceps	1	4	1	2	6	3	3	2	0	5
cacajao_calvus	4	3	4	2	5	0	2	1	3	3
cebuella_pygmea	5	1	3	3	1	2	1	2	6	2
cebus_capucinus	1	2	1	1	1	3	4	6	5	4
erythrocebus_patas	3	2	4	4	1	4	4	2	2	2
macaca_fuscata	4	4	2	3	4	3	3	1	4	2
mico_argentatus	0	4	4	2	1	3	4	3	2	3
rachypithecus_johnii	3	1	1	4	2	2	6	2	1	4
saimiri_sciureus	2	3	2	3	3	5	2	5	3	0

Figure 4.2: visualization of the confusion matrix of the model performance

With Area Under the Curve (AUC) values ranging from 0.39 to 0.55, the ROC curve graph shows that the classification performance of the model across 10 different classes is generally low as shown in figure 4.3. The majority of AUC values are in the neighbourhood of 0.5, meaning that the class distinction performance of the model is just slightly superior to random guessing. Class 9 has the lowest AUC (0.39), indicating severe difficulties in properly detecting occurrences of this class, whereas Class 7 displays the highest AUC (0.55), suggesting significantly better discriminating. Most ROC curves closely resemble the random guess baseline, indicating that the discriminatory capacity of the model is restricted.

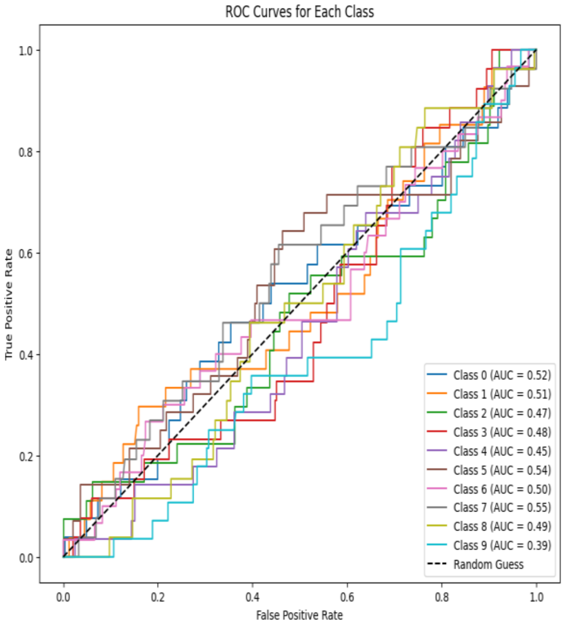


Figure 4.3: visualisation of AUC of the model performance on the different classes of monkeys

Despite reaching and sustaining 100% training accuracy by epoch 2, the model seems to have had trouble generalising due to variations in validation accuracy (ranging from 62.5 to 93.75%) and higher validation loss at specific periods.

4.1 Comparison with Other Works in Literature

Our results are consistent with the findings of Brust et al. (2017), who obtained an accuracy of 80.3% for identifying individual gorillas using a deep learning pipeline that integrated the YOLO model and AlexNet, compared to other studies in the field. Similarly, Freytag et al. (2016) reported a recognition accuracy of 91.99% for chimpanzee identification using CNNs. However, this method was also susceptible to challenges associated with overfitting and generalising diverse datasets. Zeng (2021) observed that their 2D CNN model achieved a high test accuracy of 96.67% for snub-nosed monkey classification. However, their study also reveals the limitations of standard CNN architectures in managing fine-grained visual differences between similar species and varying environmental conditions. The Inception V3 model achieved a peak validation accuracy of 93.75% in our study, which is a competitive performance. However, the observed fluctuations in validation metrics suggest that additional improvements are necessary to achieve robust generalisation, as indicated in the studies above. These comparisons underscore the fact that, even though deep learning models, such as CNNs or the Inception V3 used in our study, can achieve high accuracy in specific contexts, additional refinements, such as the integration of multi-task learning or attention mechanisms, are frequently required to manage complex, real-world scenarios more effectively (He et al., 2015; Szegedy et al., 2016).

The model exhibits potential for deployment in real-time monitoring systems, as evidenced by its consistent training accuracy and maximal validation accuracy of 93.75%.

5.0 Conclusion

This research shows how to implement a transfer learning-based deep CNN candidate called the Inception V3 model to categorise and classify monkey species using just an image dataset. Our findings demonstrate that the model can achieve impressive levels of accuracy during training, with 100% achieved in a short time and a peak validation accuracy of 93.75%. The inconsistent validation accuracy and loss numbers show problems with generalising to new data, mostly due to overfitting. Data augmentation measures, including colour jittering, random rotations, and flips, helped decrease overfitting and improve data diversity. Our results are competitive with other research, including those by Brust et al. (2017) and Freytag et al. (2016), and they support the use of deep learning models in wildlife monitoring and conservation applications. However, similar to numerous other deep learning applications, our model's efficacy is restricted by the calibre and variety of the training data and the requirement for regularisation strategies to reduce overfitting. Future research should investigate more sophisticated techniques like multi-task learning, attention processes, and synthetic data creation, utilising Generative Adversarial Networks (GANs) to enhance model robustness and generalisation capabilities further.

The Inception V3 model offers a promising framework for automating monkey species identification and classification. This framework could greatly improve conservation efforts by allowing effective, real-time monitoring in difficult and demanding conditions. To ensure this model's efficacy across various environmental circumstances and datasets, overcoming the overfitting and generalisation restrictions will be imperative before implementing it in realistic real-world applications. Further model validation and improvement will open the door to more dependable and scalable wildlife conservation solutions.

Future Directions and Recommendations:

Future work should prioritise improving the generalisability of the model by rebalancing the dataset, specifically by fixing any class imbalances. To lessen the likelihood of overfitting and obtain more representative training samples, researchers should investigate other data augmentation methods such synthetic data generation and the application of Generative Adversarial Networks (GANs). To further stabilise the training process, various learning rates or use adaptive learning rate schedules could be incorporated.

References

[1] Aldi, F., Nozomi, I., Sentosa, R. B., & Junaidi, A. (2023). *Machine Learning to Identify Monkey Pox Disease*. *8*(3), 1335–1347.

[2] Amran, A. A., On, C. K., Ariffin, S., Karim, A., Hung, L. P., See, C. S., Simon, D., Rossdy, M., & Jing, C. (2025). *Bornean Orangutan Nest Classification using Image Enhancement with Convolutional Neural Network and Kernel Multi Support Vector Machine Classifier*. *2*(2), 187–204.

[3] Brust, C., Burghardt, T., Groenenberg, M., Christoph, K., Group, C. V., & Schiller, F. (n.d.). *Towards Automated Visual Monitoring of Individual Gorillas in the Wild*. 2820–2830.

[4] Clapham, M., Miller, E., Nguyen, M., & Horn, R. C. Van. (2022). Multispecies facial detection for individual identification of wildlife : a case study across ursids. *Mammalian Biology*, *102*(3), 943–955. https://doi.org/10.1007/s42991-021-00168-5

[5] Freytag, A., Rodner, E., Simon, M., Loos, A., & Hjalmar, S. K. (2016). *Chimpanzee Faces in the Wild : Log-Euclidean CNNs for Predicting Identities and Attributes of Primates*.

[6] Guan, Y., Lei, Y., Zhu, Y., Li, T., & Xiang, Y. (2023). Face recognition of a Lorisidae species based on computer vision. *Global Ecology and Conservation*, *45*(May), e02511. https://doi.org/10.1016/j.gecco.2023.e02511

[7] Kamepalli, S. (n.d.). *Animal Breed Classification and Prediction Using Convolutional Neural Network Primates as a Case Study*.

[8] Kimly, Y., Lany, M., Soy, V., & Kor, S. (2023). *Animal Classification using Convolutional Neural Network*. *December*.

[9] Krizhevsky, A., & Hinton, G. E. (n.d.). *ImageNet Classification with Deep Convolutional Neural Networks*. 1–9.

[10] Lin, C., Li, L., Luo, W., & Wang, K. C. P. (2019). *Transfer Learning Based Traffic Sign Recognition Using Inception-v3 Model*. 242–250.

[11] Maduwuba, E., & Kamal, D. R. (n.d.). *Performing Image Classification for 10 Different Monkey Species using CNN By*. 1–7.

[12] Miao, Z., Gaynor, K. M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M. S., Mcinturff, A., Bowie, R. C. K., Nathan, R., Yu, S. X., & Getz, W. M. (2019). Insights and approaches using deep learning to classify wildlife. *Scientific Reports*, *November 2018*, 1–9. https://doi.org/10.1038/s41598-019-

44565-w

[13] Nusantara, B. (2023). *Detection of endangered species proboscis monkey using computer vision technique on low compute device*. 1–21.

[14] Pawara, P., Okafor, E., Groefsema, M., He, S., Schomaker, L. R. B., & Wiering, M. A. (2020). *One-vs-One classification for deep neural networks*. *108*. https://doi.org/10.1016/j.patcog.2020.107528

[15] Pellegrini, T., Pellegrini, T., African, D., & Pellegrini, T. (2021). *Deep-learning-based central African primate species classification with MixUp and SpecAugment To cite this version : HAL Id : hal-03298384 Deep-learning-based central African primate species classification with MixUp and SpecAugment*.

[16] Rajalingham, X. R., Issa, X. E. B., Bashivan, X. P., Kar, X. K., Schmidt, K., & Dicarlo, X. J. J. (2018). *Large-Scale , High-Resolution Comparison of the Core Visual Object Recognition Behavior of Humans , Monkeys , and State-of-the-Art Deep Artificial Neural Networks*. *38*(33), 7255–7269. https://doi.org/10.1523/JNEUROSCI.0388-18.2018

[17] Recognition, I. (2022). *Adaptive Modular Convolutional Neural Network for*.

[18] Schofield, D., Nagrani, A., Zisserman, A., Hayashi, M., & Matsuzawa, T. (2019). *Chimpanzee face recognition from videos in the wild using deep learning*. *September*, 1–9.

[19] Shi, Y., Tymoszek, C., & Jain, A. K. (2018). *Face Recognition : Primates in the Wild Face Recognition : Primates in the Wild ∗*. *October*. https://doi.org/10.1109/BTAS.2018.8698538

[20] Shukla, A., Cheema, G. S., Anand, S., Qureshi, Q., & Jhala, Y. (1878). *Primate Face Identification in the Wild*. 1–13.

[21] Trapanotto, M., Nanni, L., & Brahnam, S. (2022). *Convolutional Neural Networks for the Identification of African Lions from Individual Vocalizations*.

[22] Wang, C., Chen, D., Hao, L. I. N., Liu, X., Zeng, Y. U., Chen, J., & Zhang, G. (2019). Pulmonary Image Classification Based on Inception-v3 Transfer Learning Model. *IEEE Access*, *7*, 146533–146541. https://doi.org/10.1109/ACCESS.2019.2946000

[23] Wiranda, N., & Putro, A. E. (2022). *Mobile-based Primate Image Recognition using CNN*. *16*(2), 149–158. https://doi.org/10.22146/ijccs.65640

[24] Xu, W., Zhang, X., Yao, L., Xue, W., & Wei, B. (2020). *A Multi-view CNN-based Acoustic Classification System for Automatic Animal Species Identification*.

[25] Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (n.d.). *How transferable are features in deep neural networks ?* 1–9.

[26] Zhang, Z. (2018). *Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels*. *NeurIPS*.

[27] Zhao, K., Feng, C., Han, G., & Yan, S. (n.d.). *Research on Similar Animal Classification Based on CNN Algorithm Research on Similar Animal Classification Based on CNN Algorithm*. https://doi.org/10.1088/1742-6596/2132/1/012001

[28] Zini, S., Gomez-villa, A., Buzzelli, M., & Bicocca, M. (n.d.). *Planckian Jitter : countering the color-crippling effects of color jitter on self-supervised training*. *1*, 1–15.

Brust, C. A., Burghardt, T., Groenenberg, M., Kuhl, H. S., & Kühl, H. S. (2017). Towards Automated Visual Monitoring of Individual Gorillas in the Wild. In Proceedings of the IEEE International Conference on Computer Vision Workshops (pp. 2820-2830). doi:10.1109/ICCVW.2017.331

[29] Freytag, A., Rodner, E., Simon, M., Loos, A., & Denzler, J. (2016). Chimpanzee Faces in the Wild: Log-Euclidean CNNs for Predicting Attributes of Human Faces. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 3420-3428). doi:10.1109/CVPR.2016.372

[30] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770-778). doi:10.1109/CVPR.2016.90

[31] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2818-2826). doi:10.1109/CVPR.2016.308

[32] Zeng, L. (2021). Animal Species Classification Using Convolutional Neural Networks. Journal of Physics: Conference Series, 2132(1), 012001. doi:10.1088/1742-6596/2132/1/012001

[33] Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

[34] Prechelt, L. (1998). Early Stopping - But When? In Neural Networks: Tricks of the Trade (pp. 55-69). Springer, Berlin, Heidelberg. doi:10.1007/3-540-49430-8_3

[35] Powers, D. M. W. (2011). Evaluation: From Precision, Recall, and F-Measure to ROC, Informedness, Markedness & Correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

[36] Shorten, C., & Khoshgoftaar, T. M. (2019). A Survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(60). doi:10.1186/s40537-019-0197-0

[37] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2020). Densely Connected Convolutional Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/CVPR.2017.243

[38] Kumar, S., Singh, R., & Rai, P. (2021). IoT-based Animal Tracking System for Smart Monitoring. IEEE Sensors Journal, 21(1), 222-233. doi:10.1109/JSEN.2020.3030561

[39] Amran, A. A., On, C. K., Ariffin, S., Karim, A., Hung, L. P., See, C. S., Simon, D., Rossdy, M.,& Jing, C. (2025). Bornean Orangutan Nest Classification using Image Enhancement with Convolutional Neural Network and Kernel Multi Support Vector Machine Classifier. 2(2), 187–204.