**Abstract**: This paper explores the application of deep learning to biodiversity monitoring by using image classification models to detect and identify wildlife species. Building on a project proposal for image-based classification, we implement and evaluate machine learning models, including Convolutional Neural Networks (CNNs), transfer learning, K-Nearest Neighbors (KNN), and Random Forests, on datasets of primates and over 500 bird species. Our aim is to examine the utility of machine learning models in estimating species abundance, particularly for elusive or rarely seen species, under varying conditions of data availability. Results demonstrate that transfer learning achieves over 84% accuracy for bird classification and that performance varies significantly by species, highlighting the potential and limitations of automated ecological monitoring tools.

**1. Introduction**

Camera traps have revolutionized biodiversity monitoring by enabling passive data collection across large spatial and temporal scales (Norouzzadeh et al., 2018; Weinstein, 2018). However, the sheer volume of imagery captured by these devices presents a new challenge: manual image classification is labor-intensive and prone to human error. To address this, researchers have increasingly turned to artificial intelligence (AI) and machine learning (ML) approaches, particularly deep learning, for automating wildlife identification.

Deep learning models, especially CNNs have shown remarkable success in image classification tasks, thanks to their ability to learn hierarchical features from raw pixel data (LeCun et al., 2015; Rawat & Wang, 2017). These architectures are particularly well-suited to ecological applications, where species may vary significantly in size, shape, posture, and background conditions.

In this study, we explore the application of deep learning for classifying wildlife species across two scenarios: one involving a low-data setting with a limited number of monkey species and another involving a large-scale bird dataset with over 500 species. By comparing custom CNNs, pretrained transfer models, and traditional classifiers such as K-Nearest Neighbors and Random Forest, we aim to assess their utility for abundance estimation and conservation monitoring under both sparse and moderate data regimes.

We also incorporate techniques such as image augmentation and sensitivity-specificity analysis to evaluate model robustness. Our goal is not only to benchmark performance but to inform best practices for integrating AI into conservation workflows.

In this project, we aim to assess the potential of deep learning in comparison to traditional machine learning algorithms to support ecological monitoring in two main scenarios:

1. Low-data environments, where only a few images exist for rare or elusive species
2. High-class diversity environments, where hundreds of species must be distinguished

Our goals are twofold:

* Evaluate the efficacy of different machine learning models (CNN, transfer learning, K-Nearest Neighbors, and Random Forest) for wildlife classification
* Investigate how these models perform in both data-scarce and data-rich settings to understand their suitability for abundance estimation tasks

By testing these models on datasets of primates (limited training data) and birds (over 500 species), we aim to provide practical insights into the strengths and limitations of AI-based biodiversity monitoring.

**2. Methodology**

We implemented a range of machine learning strategies to perform wildlife species classification using image data. These included a custom-built Convolutional Neural Network (CNN), transfer learning with pretrained models (Xception and MobileNetV2), and traditional classifiers such as K-Nearest Neighbors (KNN) and Random Forest. Each model was applied to two datasets of varying size and complexity: a low-data primate dataset with approximately ten images per species, and a larger bird dataset containing over 500 species with around 100 images each. The following sections detail the architecture design, feature extraction, data augmentation procedures, and model evaluation protocols used in this study.

**2.1 Custom Convolutional Neural Network (CNN)**

A diagram of a red and blue square

AI-generated content may be incorrect.

*Figure 1. Depiction of a CNN*

To begin our evaluation, we construct a custom Convolutional Neural Network (CNN) from scratch. This model serves as a baseline to explore how well a manually designed architecture can perform in classifying wildlife species when training data is sparse, an important condition for monitoring rare or elusive animals such as primates that may only appear in a few camera trap images per year.

The architecture includes stacked convolutional layers with increasing filter sizes to learn hierarchical image features (Figure 1). Formally, each convolutional operation in the network is expressed as:

Where:

is the output feature at location in layer

is the convolutional kernel (or filter)

is the input feature from the previous layer

is the bias term

is the activation function, typically ReLU:

**2.2 Transfer Learning with Xception**

To address the challenge of limited labeled data, we turn to a technique called transfer learning (Figure 2). Instead of training a model from scratch, we begin with a model called Xception that has already been trained on a large, general image dataset (ImageNet). This pre-trained model has learned how to detect basic features like edges, shapes, and textures, skills that are useful for recognizing all kinds of objects, including animals.

We use Xception as a feature extractor, meaning we freeze its convolutional base so it doesn’t relearn anything, and simply use the patterns it already knows to interpret our wildlife images. On top of this frozen base, we add a new classification head (a few dense layers) that is trained specifically to recognize the species in our dataset.

Technically, Xception's convolutional architecture is made more efficient through the use of depthwise separable convolutions, which break the convolutional process into two steps:

* Depthwise convolution applies a single filter to each input channel separately.
* Pointwise convolution then combines these outputs using 1x1 convolutions.

This structure reduces computational cost and improves learning efficiency.

In our project, this method was especially powerful for classifying bird species, where we had roughly 100 images per class. Even without thousands of training images, Xception's pre-learned knowledge helped us achieve stronger performance. It’s like borrowing expertise from a model trained on millions of examples and applying it to a new but related problem.

A red arrow pointing to a black line

AI-generated content may be incorrect.

*Figure 2. Example Representation of Transfer Learning Process*

2.3 Traditional Models: Random Forest and K-Nearest Neighbors

Random Forest is an ensemble method where predictions are made by aggregating outputs from multiple decision trees:

Each tree votes on the predicted label, and the label with the most votes becomes the model’s final prediction. Random Forests are particularly good at reducing variance and overfitting in structured datasets, especially when feature inputs are well-defined and relatively low-dimensional.

K-Nearest Neighbors is a simple instance-based learner that assigns a class label to a sample based on the majority class among its closest neighbors in feature space. The distance between two points is typically computed using the Euclidean distance formula:

Where is the input sample, is a training point, and is the feature dimension.

However, in the context of our project, both Random Forest and K-Nearest Neighbors faced serious challenges:

* Feature Representation: Unlike CNNs, these traditional models cannot automatically learn spatial hierarchies (like edges, textures, or shapes) from image data. They rely entirely on extracted feature vectors (in our case, flattened intermediate CNN features), which may miss subtle, localized visual patterns critical for species identification.
* High-Dimensionality and Limited Data: The wildlife image data, especially the monkey dataset with very few examples per class, created a "curse of dimensionality" problem for K-Nearest Neighbors, and Random Forests were prone to overfitting noisy feature representations without the ability to extract more meaningful patterns.
* Fine-Grained Differences: In wildlife images, many species differ by very small visual cues (e.g., fur color, slight differences in size or beak shape). Deep learning models like CNNs can learn these fine-grained distinctions, whereas traditional methods often treat these differences as noise.

These models served as a point of comparison to evaluate whether simpler approaches could provide acceptable performance, particularly when computational resources are limited or deep learning models are impractical to deploy in field settings.

**2.4 Image Augmentation**

To mitigate overfitting and help the model generalize beyond the limited training examples, we applied a series of image augmentation techniques during training. These transformations artificially increase the diversity of the dataset by simulating different environmental conditions, camera angles, and object appearances (Figure 3). Specifically, we applied:

* Rotation (small random angles) to mimic different animal poses and camera placements.
* Horizontal and vertical flipping to simulate left-right symmetry and different animal orientations.
* Zooming (both in and out) to vary the scale of objects in the frame.
* Brightness and contrast adjustments to account for different lighting conditions, such as daytime versus dusk imagery.

By presenting the model with varied versions of the same images, augmentation forces it to learn robust, generalizable patterns (e.g., the shape of a bird’s wing or the outline of a monkey’s face) rather than memorizing specific pixel arrangements.

This was particularly important for our study, where some species were represented by only a handful of training images. Without augmentation, models would likely have overfit to these few samples and performed poorly on new, unseen images, especially under different field conditions common in camera trap datasets.

Thus, image augmentation was an important step to enhance model resilience and simulate the diversity encountered in real-world wildlife monitoring applications.

A collage of a lion

AI-generated content may be incorrect.

*Figure 3. Augmenting a Single Image for Training*

**3. Results**

**3.1 Model Performance and Evaluation (Monkey Data)**

The custom CNN achieved a training accuracy of approximately 84% and a validation accuracy of around 71% after 200 training epochs. While this demonstrates effective learning on the training set, the widening gap between the training and validation curves (Figure 4) suggests overfitting, likely due to the limited amount of training data per species, especially in the monkey dataset.

A graph showing the different types of graphs

AI-generated content may be incorrect.

*Figure 4. Training/Validation Model Evaluation (Accuracy and Loss for Custom CNN)*

These plots highlight the model’s learning trajectory: accuracy increases over time while the loss steadily decreases. However, after around 125 epochs, validation accuracy begins to plateau and even fluctuate, indicating that the model has likely memorized training examples and may struggle with generalization.

To address these limitations, we implemented a transfer learning approach using the Xception model, which was pre-trained on the ImageNet dataset. This method demonstrated excellent convergence behavior, achieving high training accuracy and rapidly decreasing loss within a small number of epochs (see Figure 5). However, when applied to the monkey dataset, the Xception-based transfer model achieved only 10% validation accuracy, underscoring the difficulty of adapting pre-trained features to extremely small and visually subtle datasets.

A graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

AI-generated content may be incorrect.

*Figure 5: Model performance across 15 epochs using Xception. The model achieves high training accuracy (~99%) and low validation loss (~0.10) on bird data; however, performance on monkey species remained constrained.*

In the monkey species classification task, we evaluated several machine learning models trained on features extracted from the Xception model. Given the small sample size, approximately ten images per species, and the visual similarity across categories, classification proved extremely challenging. The custom CNN achieved a test accuracy of 12%, slightly outperforming the transfer learning model, which reached 10%. Despite the theoretical advantages of CNNs in capturing spatial and hierarchical patterns, both deep learning approaches struggled to generalize due to the extremely limited training data, leading to significant overfitting and poor performance on unseen examples.

Surprisingly, traditional machine learning models outperformed these deep learning methods. Random Forests achieved 35% accuracy, and K-Nearest Neighbors reached 21%. Although these models are typically less suited for image data because they do not capture spatial relationships, they were able to make better use of extracted features under conditions of extreme data scarcity. Most notably, a simple feedforward neural network trained on the same feature set achieved the highest performance, with a test accuracy of 73%. This result was unexpected, as fully connected networks generally lack the spatial modeling capabilities of CNNs; however, in this case, the lower model complexity appears to have better matched the small sample size, reducing the risk of overfitting.

These findings underscore an important lesson for ecological image classification tasks: model complexity must be carefully matched to the size and quality of the available data. In cases where only limited labeled examples exist, lightweight neural networks or traditional machine learning methods trained on strong feature representations may outperform deeper architectures. For conservation monitoring scenarios where data on rare or elusive species are often sparse, leveraging extracted features with simpler models could offer a more practical and effective solution than training large convolutional networks from scratch.

**3.2 Overall Model Comparison and Key Takeaways (Monkey Data)**

Table 1 summarizes the final test accuracies achieved by all models evaluated on the monkey dataset. Despite the typical advantage of convolutional architectures for image data, a simple fully connected neural network surprisingly achieved the highest accuracy at 73%, highlighting the critical role of matching model complexity to dataset size and feature representation.

| **Model** | **Final Accuracy (%)** |
| --- | --- |
| Fully Connected Neural Network | 73% |
| Random Forest | 35% |
| K-Nearest Neighbors (KNN) | 21% |
| Custom CNN | 12% |
| Transfer Learning (Xception) | 10% |

*Table 1: Model Performance on Monkey Dataset*

Across models, we observed several important trends. Fully connected networks benefited from deep feature embeddings extracted via Xception, learning flexible non-linear patterns without the risk of overfitting a small number of samples. Random Forest and K-Nearest Neighbors classifiers captured some structure but lacked the capacity to model subtle visual differences between species. Custom CNNs and transfer learning approaches, although effective during training, suffered from severe generalization failure when evaluated on unseen data.

These results emphasize that even advanced techniques like transfer learning can be ineffective when sample sizes are extremely small or when the domain of the pre-trained model (e.g., ImageNet) diverges from the target dataset (wild monkeys). Traditional machine learning models, though weaker for spatially complex data, may outperform deep networks when paired with strong feature extraction and low sample regimes.

From a conservation perspective, this suggests that abundance estimation models must be carefully calibrated to the available data. In scenarios where rare species yield very few observations, lightweight classifiers trained on pre-extracted deep features may offer more reliable and practical solutions than large, high-capacity models prone to overfitting.

**3.3 Sensitivity, Specificity, and Feature Interpretations (Monkey Data)**

To deepen our evaluation, we also computed sensitivity and specificity metrics for each species class (Figure 6). Sensitivity values, which reflect true positive rates, were relatively low across classes, ranging from approximately 3.7% to 15%, while specificity values were consistently high, around 90%. This imbalance suggests that while the models are effective at recognizing when a species is not present, they struggle to correctly identify true instances of species occurrence, a common challenge in ecological monitoring where positive observations are rare relative to background images.



*Figure 6. Heatmap of Evaluation Metrics for Custom CNN*

To further interpret model behavior, we visualized activation maps from early and mid-level convolutional layers of the Xception model. The early layers (e.g., block1\_conv1, Figure 7) predominantly captured broad structures such as contours, fur textures, and head outlines. In contrast, mid-level layers (e.g., block3\_sepconv1, Figure 8) learned more complex patterns, including facial structures and finer textures.A collage of different animals

AI-generated content may be incorrect.

*Figure 7.* ***Early layers*** *capture broad edges, textures, and low-level shapes. As seen in `block1\_conv1`, these features often highlight the overall contour of the monkey, such as its silhouette, fur texture, and head outline.*

A collage of squares with images of a bear

AI-generated content may be incorrect. *Figure 8.* ***Mid-level layers*** *begin to focus on more abstract patterns and finer-grained structures. In `block3\_sepconv1`, we can see more nuanced activations—possibly corresponding to facial structure, posture, or background texture.*

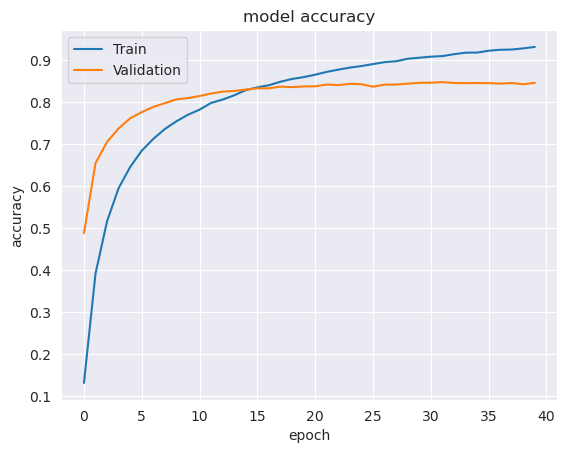
These visualizations confirm that the transfer learning model learned semantically meaningful features despite overall classification struggles. However, they also highlight why subtle differences between visually similar species remained difficult to distinguish.

**3.4. Bird Species Classification at Scale**

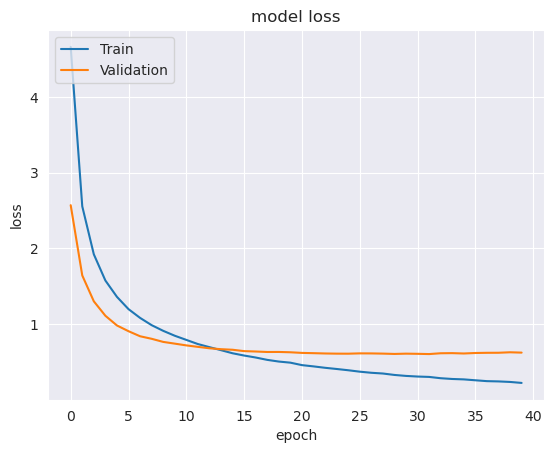
To extend our investigation into high-diversity ecological datasets, we developed and evaluated another custom CNN model designed to classify over 500 distinct bird species. This task posed a unique challenge due to the sheer number of target classes, yet it also presented an opportunity to assess model scalability and generalizability in a more data-rich environment compared to our prior low-data primate classification task.

We utilized MobileNetV2 as the base architecture for transfer learning. The model was initialized with weights pretrained on the ImageNet dataset and customized with dense layers, dropout regularization, and a final softmax output layer for 500-class classification. The model was compiled using the Adam optimizer, which provided efficient, adaptive learning rates during training. To prevent overfitting and ensure model robustness, we implemented several callbacks including early stopping, which monitored validation loss and terminated training if no improvement was observed over eight consecutive epochs, and model checkpointing, which saved the weights corresponding to the best validation accuracy.

Over the course of 40 training epochs, the model demonstrated consistent and significant improvements in performance. Training accuracy increased from 13% to over 93%, while validation accuracy rose steadily and plateaued around 84.5% (Figure 9). Concurrently, validation loss declined from 2.57 in the first epoch to just 0.62 by the final epoch, indicating successful convergence and minimal overfitting. These results confirm that MobileNetV2, when combined with appropriate regularization and augmentation techniques, can serve as an effective model for fine-grained bird species classification.



*Figure 9. CNN Evaluation (Accuracy) for Bird Data*



*Figure 10. CNN Evaluation (Loss) for Bird Data*

Following training, the model was evaluated on an unseen test set to measure its generalization capability. The model achieved a test accuracy of 84.11% and a test loss of approximately 0.605 (Figure 10). These metrics reinforce the conclusion that the model was not only capable of learning species-specific features from the training data but also maintained this performance when classifying images outside of the training and validation sets.

The close alignment between training and validation accuracy curves suggests that the model maintained a strong generalization capability throughout training. The loss curves further support this conclusion, with both training and validation losses decreasing in parallel. Importantly, the narrow gap between the two metrics implies a well-regularized and robust model that can be deployed confidently in field settings.

The implications of these results for avian biodiversity monitoring are substantial. A classification model capable of accurately identifying hundreds of bird species can automate species labeling across vast datasets collected from camera traps, drones, or citizen science platforms. This automation can significantly reduce the labor burden on researchers, enabling more rapid biodiversity assessments and supporting real-time conservation decision-making. Moreover, the model’s strong generalization ability means it could feasibly be embedded in low-power or edge computing devices for deployment in remote field environments, contributing to real-time monitoring and alert systems.

Furthermore, the model's utility extends beyond species identification. By combining predictions with geographic and temporal metadata, it becomes possible to perform automated abundance estimation, build dynamic species distribution models, and detect shifts in species presence over time, functions critical for monitoring climate change impacts and habitat degradation. This experiment demonstrates the feasibility and value of deep learning for supporting scalable, automated, and accurate avian biodiversity surveillance.

**3.5 Visualizing Predictions on Test Data (Bird Data)**

To further evaluate the generalization ability of our trained model, we applied it to the unseen test set and visualized a random sample of 25 predictions. Each image is annotated with the true and predicted species label (Figure 11).

* Correct predictions are displayed in green (True = Predicted).
* Incorrect predictions are highlighted in red, indicating misclassifications.

A collage of different birds

AI-generated content may be incorrect.

*Figure 11. 25 Predicted Bird Images (majority classified correctly)*

The model correctly identified many species, including Ovenbird, Mallard Duck, and Crested Nuthatch, with high confidence and visual clarity (Figure 11). These correct classifications demonstrate that the model has effectively internalized key visual features (e.g., plumage color, beak shape, posture) necessary for discriminating between species.

Misclassifications, such as predicting Fire Tailed Myzornis instead of Red Legged Honeycreeper, or Cinnamon Flycatcher in place of Flame Tanager, typically involved birds with similar body structures or overlapping coloration occasionally (Figure 11). These errors suggest that while the model performs well in general, it struggles with subtle intra-class variations or inter-class visual overlap, common challenges in fine-grained image classification.

Lighting conditions, image quality, and background clutter may have also contributed to misclassification. For example, shadows or occlusions could have altered the appearance of key identifying features, while background textures may have distracted the model’s learned filters.

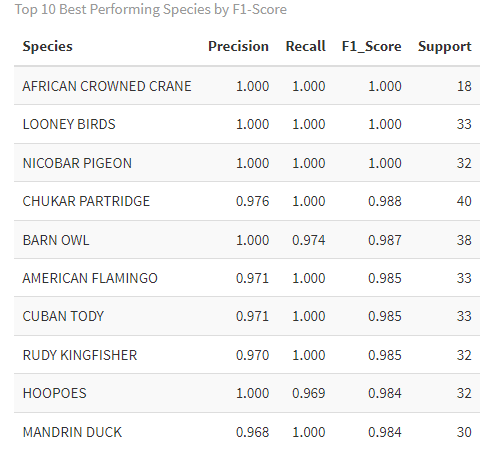
These prediction results reaffirm the model’s potential for use in real-world biodiversity workflows. In practical applications:

* Correct classifications can dramatically reduce the burden on human annotators, enabling faster, more scalable ecological monitoring.
* Even partial correctness (e.g., top-2 or top-5 accuracy) could still be highly useful in semi-automated labeling systems or for narrowing down potential species for expert validation.
* Misclassifications pinpoint areas for improvement, highlighting the value of expanding the dataset with additional training images, improving image augmentation strategies, or refining the model architecture.

Ultimately, this section demonstrates that even with a relatively compact CNN model and 100 training images per class, high-accuracy species predictions are achievable. The ability to automate bird classification at this scale opens the door for large-scale applications in conservation biology, such as longitudinal abundance tracking, migration pattern analysis, and early warning systems for declining species.

**3.6 Classification Summary: Best vs. Worst Performing Species (Bird Data)**

To better understand how the model performs across individual bird species, we generated a full classification report. This allowed us to sort and visualize the top and bottom performing classes based on F1-score, which balances precision and recall (see Figures 12 & 13 below).



*Figure 12. Top 10 Performing Species by F1-Score for CNN*



*Figure 13. Bottom (difficult to predict and classify) 10 Performing Species by F1-Score for CNN*

Species such as African Crowned Crane, Looney Birds, and Nicobar Pigeon achieved perfect F1-scores of 1.0, indicating consistent and accurate classification. Other top-performing birds included Barn Owl, American Flamingo, and Hoopoes, each with F1-scores well above 0.98.

In contrast, the lowest performing classes, such as Gilded Flicker, Northern Flicker, and Antbird had F1-scores below 0.50. These species may be more difficult to distinguish visually or were represented with lower-quality or more variable images. These findings offer valuable guidance on where future data collection or model refinement could be focused.

This analysis further underscores the importance of per-class evaluation in biodiversity applications, especially when some species are rare, endangered, or visually similar to others. It helps identify species for which additional training data may be needed, and it allows conservationists to weigh model outputs appropriately depending on their confidence and ecological priorities.

**4. Discussion**

This study highlights the strengths and limitations of deep learning approaches for automated species classification in biodiversity monitoring. Across both low-data (monkey) and high-diversity (bird) settings, model performance was strongly influenced by training data volume, visual complexity, and the choice of modeling architecture (LeCun et al., 2015).

In the low-data monkey dataset, both the custom convolutional neural network (CNN) and the transfer learning model (Xception) struggled to generalize well, achieving validation accuracies of only 12% and 10%, respectively. Despite using image augmentation and feature extraction, these deep models suffered from overfitting due to the extremely limited number of training examples per class. Interestingly, simpler approaches performed better: a fully connected neural network trained on Xception-extracted features achieved 73% accuracy, while Random Forests and K-Nearest Neighbors reached 35% and 21% accuracy, respectively. This was surprising, as traditional machine learning models typically do not excel in image classification tasks where spatial hierarchies matter. However, in scenarios of extreme data scarcity, their lower model complexity appeared to match the data conditions better, avoiding the overfitting seen in deeper architectures.

Sensitivity and specificity analyses provided further insight. Sensitivity (true positive rate) remained low across species classes (ranging from 3.7% to 15%), indicating that the models frequently missed actual species presences. Specificity, on the other hand, remained consistently high (~90%), meaning models were better at recognizing absences. This imbalance reflects a common challenge in ecological datasets, where positive examples are rare relative to background images (Beery et al., 2019).

Feature visualizations of the Xception model confirmed that meaningful patterns, such as contours, textures, and facial structures, were learned at both early and mid-level convolutional layers. However, given the subtle visual differences between monkey species and the small number of training images, even strong feature extraction was insufficient to guarantee reliable classification.

In contrast, the bird dataset demonstrated the power of deep learning when moderate amounts of training data (~100 images per class) were available. The transfer learning model (MobileNetV2) achieved a test accuracy of 84%, showing rapid convergence and minimal overfitting. This validates the use of pretrained convolutional backbones for large-scale species classification tasks (Howard et al., 2017). Nonetheless, species-level analysis revealed variation: some birds, such as the African Crowned Crane and Nicobar Pigeon, were classified with perfect accuracy, while others like the Gilded Flicker and Antbird remained more difficult to distinguish. These challenges likely stem from inter-class visual similarities and variable image quality (Kellenberger et al., 2020).

Finally, the comparison between traditional ML models and deep neural networks reaffirmed a key principle: while shallow classifiers can sometimes outperform deep models under conditions of extreme data scarcity, fully capturing the complexity of ecological image data requires the hierarchical feature learning capabilities of deep learning. Future work should focus on improving sample diversity, exploring semi-supervised learning strategies, and developing explainable AI tools to enhance model interpretability and trust in conservation applications (Christin et al., 2019).

**5. Conclusion**

This study evaluated the effectiveness of CNNs, transfer learning, and traditional machine learning classifiers for species identification in wildlife imagery. We tested these methods across two challenging datasets: one featuring monkeys with very few images per species, and another featuring birds across 500 classes with moderate image availability.

Our results demonstrate that:

* Transfer learning significantly improved performance when moderate-sized datasets were available, achieving over 84% accuracy on the bird dataset (Tan & Le, 2019).
* Custom CNNs were effective but prone to overfitting in low-data conditions without sufficient training examples.
* Traditional classifiers like Random Forest and K-Nearest Neighbors, while typically less suited for image tasks, surprisingly outperformed deep models when trained on strong feature representations in the extreme low-data monkey scenario.
* Fully connected neural networks trained on extracted features achieved the highest accuracy (73%) in the monkey dataset, illustrating the importance of matching model complexity to data availability.

Evaluation metrics such as sensitivity, specificity, and visual inspection of feature maps revealed that even high-performing models can struggle with subtle inter-class variations and rare species detection.

From a conservation standpoint, these findings suggest that deep learning, particularly when combined with transfer learning, is a powerful tool for large-scale biodiversity monitoring. However, careful attention must be paid to data quality, sample size, and model interpretability. Future research should pursue hybrid models that integrate visual, spatial, and contextual data, leverage semi-supervised learning approaches, and develop human-in-the-loop systems to validate model outputs. Ultimately, by combining camera traps with AI, we can create scalable, accurate tools for safeguarding the world’s biodiversity (Weinstein, 2018).

References

Beery, S., Van Horn, G., & Perona, P. (2019). Recognition in terra incognita. In Proceedings of the European Conference on Computer Vision (ECCV), 456–473.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Christin, S., Hervet, É., & Lecomte, N. (2019). Applications for deep learning in ecology. Methods in Ecology and Evolution, 10(10), 1632–1644.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.

Howard, A. G., et al. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

Kellenberger, B., Marcos, D., & Tuia, D. (2020). Neural networks in wildlife research: A review. Methods in Ecology and Evolution, 11(10), 1490–1506.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Norouzzadeh, M. S., et al. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716–E5725.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation, 29(9), 2352–2449.

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432.

Tan, M., & Le, Q. V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946.

Weinstein, B. G. (2018). A computer vision for animal ecology. Journal of Animal Ecology, 87(3), 533–545.