Comparative Study of CNN Architectures for Poultry Disease Detection

Rogelio Salas Vargas
Department of Artificial Intelligence
Universidad Panamericana
Aguascalientes, Mexico
0247152@up.edu.mx

***Abstract*—The early detection of poultry diseases is crucial for preserving flock health, reducing economic losses, and ensuring high-quality poultry products. This study presents a comparative analysis of deep learning models for the automatic classification of poultry diseases using fecal and symptomatic images. Specifically, we evaluate the performance of three Convolutional Neural Network (CNN) architectures: ResNet-18, ResNet-34, and ResNet-50, trained on the Poultry Pathology Visual Dataset [7].**

The methodology includes data preprocessing, augmentation, model training, and performance evaluation based on accuracy, precision, recall, and F1-score. By comparing these architectures, we aim to identify the optimal balance between classification performance and computational efficiency.

Although deeper models have the potential to achieve higher accuracy, the results of this study show that increased depth does not always guarantee better performance. This comparative analysis highlights the importance of carefully selecting models based not only on accuracy but also on computational efficiency and practical applicability. These findings provide valuable guidance for deploying suitable models in poultry disease diagnosis, ultimately contributing to faster, more cost-effective, and accurate identification of diseases for breeders and veterinarians.

***Index Terms*—Poultry diseases, Deep learning, Model comparison, ResNet, Convolutional Neural Networks, Fecal images, Artificial Intelligence.**

I. Introduction

The early detection of poultry diseases remains a critical challenge in the poultry industry and veterinary medicine. Disease outbreaks can spread rapidly, leading to significant production losses, economic setbacks, and public health concerns. Traditional diagnostic methods rely heavily on manual observation or laboratory testing, which can be time-consuming, expensive, and require specialized expertise.

In recent years, the integration of Artificial Intelligence (AI), particularly deep learning, has opened new possibilities for automating and improving disease detection. Convolutional Neural Networks (CNNs), in particular, have demonstrated high performance in image classification tasks, making them suitable candidates for identifying poultry diseases from visual cues such as fecal images.

This study focuses on the implementation and comparison of multiple deep learning architectures—ResNet-18, ResNet-34, and ResNet-50—for classifying poultry diseases using the Poultry Pathology Visual Dataset [7]. The objective is to evaluate and contrast the performance of these models in terms of accuracy, precision, recall, and other relevant metrics to identify the most efficient and scalable solution.

Beyond achieving high accuracy, this comparative analysis aims to understand the trade-offs between model complexity and performance, which is essential for deploying AI systems in resource-constrained environments such as farms or rural clinics.

By systematically benchmarking these architectures, this work contributes to the growing field of AI-assisted diagnostics in agriculture and veterinary medicine, promoting the adoption of reliable, automated tools that can enhance disease management, reduce mortality rates, and support early interventions in poultry farming.

II. Deep Learning in Animal Health

Convolutional Neural Networks (CNNs) have demonstrated outstanding performance in medical imaging tasks such as dermatology, radiology, and ophthalmology, due to their ability to automatically extract hierarchical spatial features from raw image data. Inspired by these successes, similar deep learning methodologies have been increasingly adopted in veterinary applications, including lameness detection in cattle, parasite identification in fish, and respiratory disease diagnosis in pigs.

In poultry health, however, deep learning applications are still emerging. This study contributes to this growing field by benchmarking the performance of different CNN architectures—ResNet-18, ResNet-34, and ResNet-50—for the classification of poultry diseases using fecal images. By comparing these models, we aim to identify architectures that balance classification performance with computational efficiency, which is crucial for real-world deployment in agricultural settings.

III. Related Work

Various approaches have been proposed in recent years to automate the detection and classification of poultry diseases using deep learning and computer vision techniques.

Srivastava and Pandey [1] presented a deep learning-based model that classifies poultry diseases using symptomatic features, showcasing the potential of convolutional neural networks (CNNs) to achieve high accuracy in differentiating between various poultry conditions. The model's performance demonstrated the capability of deep learning systems to support early detection and decision-making in poultry farming.

M. Z. Degu and G. L. Simegn [2] proposed a smartphone-based detection system that identifies poultry diseases by analyzing chicken fecal images. Their work highlights the advantages of deploying lightweight deep learning models on mobile devices, enabling real-time and accessible diagnostics for farmers in remote or low-resource areas.

A. Nakrosis et al. [3] introduced a non-invasive technique based on computer vision to classify poultry health by analyzing droppings. Their system employed a CNN trained on image data of poultry feces, demonstrating a viable alternative to invasive medical diagnosis and contributing to animal welfare through stress-free monitoring.

J.-R. Bisailon et al. [4] focused on the classification of gross abnormalities observed during postmortem inspection in Canadian poultry abattoirs. Their approach relied on a hazard identification decision tree that helped categorize visual signs of disease or defects, providing a standardized method for health assessment in poultry processing environments.

M. S. Uwadia Osagie and F. Cyril-Musa [5] implemented a CNN-based classifier to categorize poultry birds based on their health state. Their results revealed that CNNs could successfully identify differences between healthy and diseased birds, supporting the implementation of automated surveillance tools in poultry farms.

Finally, H. Hemalatha et al. [6] proposed a system for real-time recognition of avian pox disease in poultry using extreme learning machines (ELM) and support vector machines (SVM) with Gaussian Radial Basis Function (GRBF). They utilized texture features extracted from images using the Gray Level Co-Occurrence Matrix (GLCM), along with statistical features such as mean, standard deviation, kurtosis, and skewness. Their findings revealed that the ELM classifier outperformed SVM in terms of accuracy, suggesting its potential for real-time poultry disease detection.

These studies collectively demonstrate the increasing relevance of deep learning models for poultry disease classification. However, most of the existing work focuses on developing a single model architecture, without conducting comparative analysis between multiple CNN models. Moreover, many are limited to small datasets or constrained platforms. In contrast, our work introduces a comparative study of ResNet-18, ResNet-34, and ResNet-50 on a large and diverse dataset of over 500,000 images, with the objective of identifying the most effective and efficient architecture for real-world deployment in poultry disease detection.

IV. Examples of Poultry Diseases in the Dataset

This section presents representative RGB images from the dataset used to train the poultry disease classification models. The images depict poultry feces, both from healthy and infected individuals, capturing a wide range of visual conditions including lighting variations, backgrounds, and visible manifestations of disease. The dataset comprises exactly 510,677 images categorized into four classes: Coccidiosis, Newcastle Disease, Salmonellosis, and Healthy. This variability introduces a complex and challenging learning environment, requiring the model to generalize across diverse visual features to achieve accurate classification.

Figures 1 through 4 display sample images corresponding to each of the four classes in the dataset.

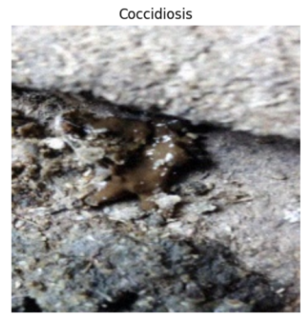

Fig. 1. Poultry with Coccidiosis

Fig. 2. Poultry with Newcastle Disease

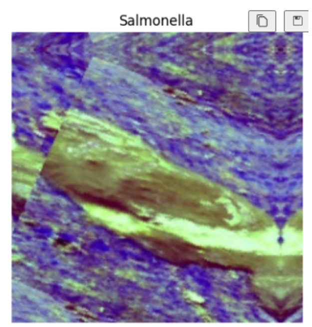

Fig. 3. Poultry with Salmonellosis

Fig. 4. Healthy Poultry

A. Data Augmentation

Data augmentation refers to a set of techniques used to artificially increase the size and diversity of training datasets, improving the generalization capabilities of deep learning models. It is especially useful in computer vision tasks, where it helps mitigate overfitting—when a model learns the training data too precisely, including noise and irrelevant details.

In this study, data augmentation was already applied to the original dataset by its creators. The dataset includes variations generated through geometric transformations such as rotations, zooming, flipping, and shifting. These pre-augmented images contribute to the model's ability to learn more robust and invariant features across different visual contexts.

Figure 5 shows examples of original images alongside their augmented versions, as provided in the dataset.

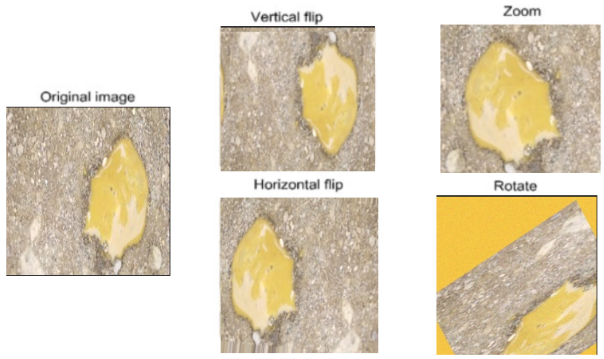

Fig. 5. Examples of original and augmented training images from the dataset.

V. Methodology

To assess the robustness and effectiveness of different convolutional neural network (CNN) architectures, a comparative study was conducted. This comparison aims to highlight the trade-offs between model complexity, training efficiency, and classification accuracy. We trained and evaluated multiple ResNet architectures under the same experimental conditions, including identical data preprocessing, batch size, learning rate, number of epochs, and optimizer settings.

Importantly, all ResNet models were initialized with pretrained weights from ImageNet, a large-scale dataset widely used in computer vision. This pretraining significantly enhances the models' effectiveness by providing a strong starting point for feature extraction, particularly in tasks with limited labeled data.

A. Model Architectures Evaluated

The selected models—ResNet-18, ResNet-34, and ResNet-50—are part of the Residual Network (ResNet) family. These architectures were chosen due to their proven performance in image classification tasks and their increasing depth and complexity:

- **ResNet-18**: A lightweight network suitable for fast training and inference with fewer parameters.
- **ResNet-34**: A deeper model offering improved representational capacity over ResNet-18.
- **ResNet-50**: A deeper network with bottleneck layers, providing higher accuracy at the cost of increased computational resources.

B. Classification Pipeline Using ResNet Architectures

The disease classification process follows a unified pipeline applied to three ResNet variants. The main stages of the pipeline include: dataset selection, image preprocessing, data loading, model definition and training, performance validation, and final evaluation.

- **Dataset Selection:**A curated dataset of bird droppings affected by different diseases is used. The dataset contains four distinct classes, each corresponding to a specific disease condition. Images are organized into folders suitable for PyTorch's `ImageFolder` structure.
- **Preprocessing:** All images are resized to 224×224 pixels and normalized using standard ImageNet parameters (`mean=[0.485, 0.456, 0.406]`, `std=[0.229, 0.224, 0.225]`). This ensures compatibility with the pretrained ResNet models.
- **Data Loading:** The `ImageFolder` class from PyTorch is used to automatically label the images. The data is divided into training, validation, and test sets. `DataLoader` is employed with a `batch_size` of 128 for efficient mini-batch processing.
- **Model:** The pipeline is applied to ResNet18, ResNet34, and ResNet50 architectures, all initialized with pretrained ImageNet weights. In each case, the final fully connected layer is modified to output four classes:

```
model.fc = nn.Linear(model.fc.in_features, 4)
```

- **Training:** The models are trained using the `CrossEntropyLoss` function and the `AdamW` optimizer. A learning rate scheduler (`ReduceLROnPlateau`) adjusts the learning rate dynamically based on validation performance. Each model is trained for 10 epochs, and the best validation result is saved as `best.pt`.

$$\text{Loss} = -\sum_{i=1}^{\substack{\text{output}\\ \text{size}}} y_i \cdot \log \hat{y}_i$$

Fig. 6. Mathematical expression of the CrossEntropyLoss function used for classification tasks. This loss measures the difference between the predicted probability distribution and the true label distribution.

- **Evaluation Criteria:** All models were trained and tested using the same dataset splits. The training configuration was consistent across all experiments, using a batch size of *X*, learning rate of *Y*, optimizer (e.g., Adam or SGD), and a fixed number of epochs. Performance was evaluated using:
  - **Accuracy**
  - **F1-Score**
  - **Training Time**
  - **Model Size**

  The following figure illustrates the end-to-end classification pipeline implemented for this study. It summarizes the key stages applied to the ResNet-18 architecture, including data preprocessing, model training, validation, and evaluation. This visual overview provides a clear understanding of the methodological workflow used in the experiments.

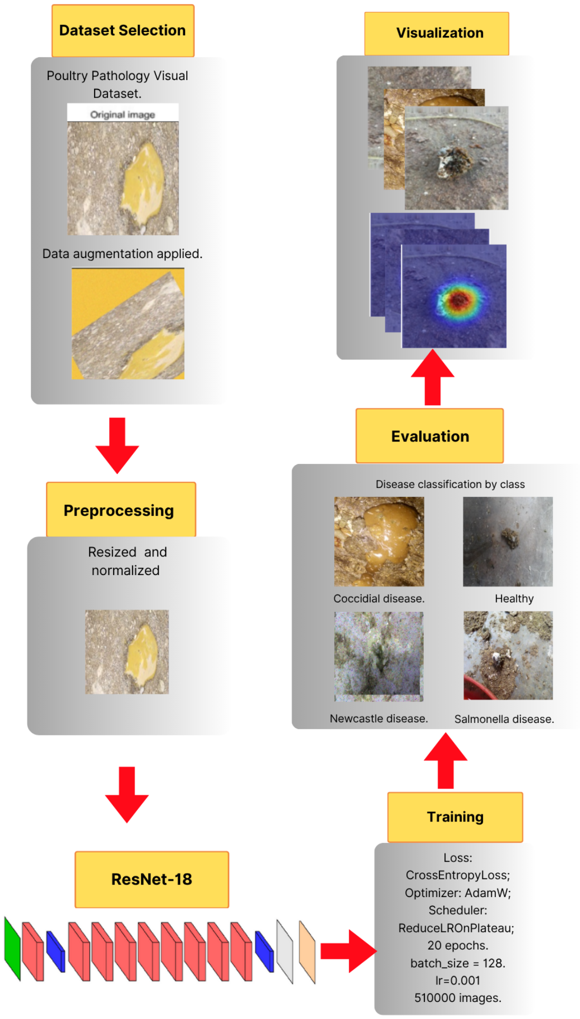

Fig. 7. Pipeline applied to ResNet-18 architecture.

C. Comparison Between ResNet Variants

Table I presents the performance of each ResNet variant:

TABLE I
PERFORMANCE COMPARISON OF RESNET ARCHITECTURES

Model	Accuracy	F1-Score	Training Time	Model Size (MB)
ResNet-18	94.17%	0.95	313 min	44
ResNet-34	93.2%	0.95	532 min	83
ResNet-50	91.5%	0.95	4615 min	98

Table I presents the performance of each ResNet variant. Among the three architectures evaluated—ResNet-18, ResNet-34, and ResNet-50—ResNet-18 achieved the highest accuracy at 94.17 percent with an F1-score of 0.95. Despite having a smaller model size (44 MB) and significantly lower training time (313 minutes), it outperformed the deeper variants. ResNet-34 demonstrated slightly lower accuracy (93.2 percent) with the same F1-score, but required more training time (532 minutes) and a larger model size (83 MB). Surprisingly, ResNet-50, the deepest and most complex model, showed the lowest accuracy (91.5 percent) while maintaining the same F1-score of 0.95. Additionally, it had the longest training time (4615 minutes) and the largest model size (98 MB). Interestingly, despite its lower complexity, ResNet-18 outperformed its deeper counterparts. This can be attributed to several factors. First, deeper models such as ResNet-50 are more prone to overfitting, especially when the added depth does not translate into learning more meaningful patterns for the given task. In our case, the visual differences among the four classes of droppings may not require deep hierarchical features, making ResNet-18 sufficiently expressive. Additionally, training was conducted on a mid-range GPU (NVIDIA GTX 1650), which may have limited the computational efficiency and convergence behavior of more complex models like ResNet-50. These findings emphasize that model simplicity can be advantageous when dealing with low inter-class variability and limited computational resources.

To further analyze the best-performing model, Table II shows the classification metrics of ResNet-18 on each class:

TABLE II
RESNET-18 CLASSIFICATION REPORT BY CLASS

Class	Precision	Recall	F1-Score	Support
Coccidiosis	0.97	0.97	0.97	18752
Healthy	0.92	0.95	0.94	17412
New Castle Disease	0.94	0.95	0.94	15888
Salmonella	0.96	0.93	0.95	18625
Accuracy			0.95	70677
Macro Avg	0.95	0.95	0.95	70677
Weighted Avg	0.95	0.95	0.95	70677

VI. MODEL EXPLAINABILITY USING GRAD-CAM

To enhance the interpretability of the model's predictions and understand which regions of the input images most influence classification decisions, we applied Gradient-weighted Class Activation Mapping (Grad-CAM). This technique produces heatmaps that visually highlight the spatial areas in an image that are most relevant to the model's output, offering insights into the reasoning behind each prediction.

Grad-CAM is particularly valuable in the context of poultry disease detection, as it allows us to validate whether the model is focusing on meaningful biological cues—such as lesions, discoloration, or abnormal textures in the feces—rather than irrelevant background information. This contributes to both transparency and trustworthiness of the system, especially in real-world agricultural applications.

Unlike other domains such as radiology, where interpretability often relies on the identification of anatomical shapes or structures (e.g., tumors or bone fractures), fecal images typically lack consistent geometric forms. Instead, discriminative features are primarily based on **color variations, texture patterns, and irregular pigmentation**. Grad-CAM enables us to verify that the model's attention aligns with these critical non-structural features, reinforcing the biological plausibility of its decisions.

Figure 8 presents Grad-CAM visualizations overlaid on example images from each of the four classes in the dataset (Coccidiosis, Newcastle Disease, Salmonellosis, and Healthy). The red areas in the heatmaps represent regions with higher importance for the classification decision, while blue areas represent low importance.

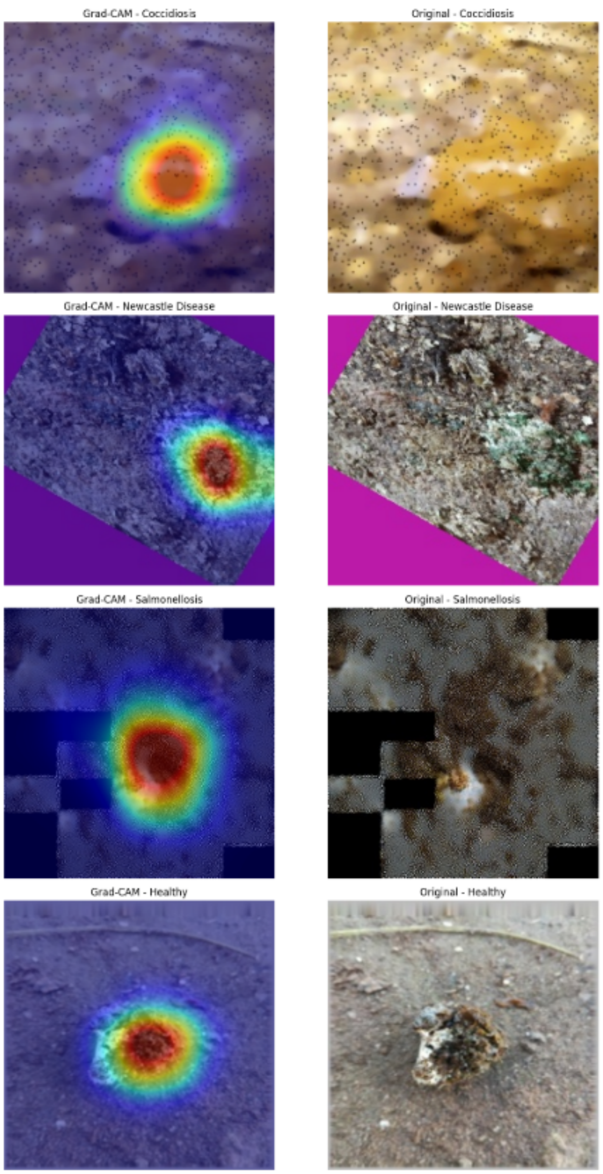

Fig. 8. Grad-CAM visualizations highlighting the regions most relevant to the model's classification decisions across different disease categories.

A. Grad-CAM Interpretation Summary

The following summarizes the model's focus based on Grad-CAM visualizations applied to fecal images from each class:

- **Coccidiosis:** High activation around reddish or stained regions in the feces. These areas often indicate the presence of blood or mucous, which are clinical signs of the disease.
- **Newcastle Disease:** The model tends to focus on regions with watery or irregular textures, which may correspond to diarrhea or abnormal digestive patterns linked to the infection.
- **Salmonellosis:** The model highlights greenish or yellowish areas in the fecal matter. These features are often associated with bile excretion or bacterial contamination.
- **Healthy:** Uniform textures and coloration throughout the image. Grad-CAM shows low or no activation, indicating no detectable abnormalities.

B. Confusion Matrix Analysis

To further evaluate the classification performance of the models, a confusion matrix is generated using the predictions on the test set. The matrix provides a detailed breakdown of correct and incorrect classifications for each of the four disease categories. Each row of the matrix represents the true class, while each column represents the predicted class. High values along the diagonal indicate good performance, as they correspond to correct predictions. Off-diagonal values represent misclassifications, helping to identify specific confusions between classes.

Next figure displays the confusion matrix for the ResNet18 model.

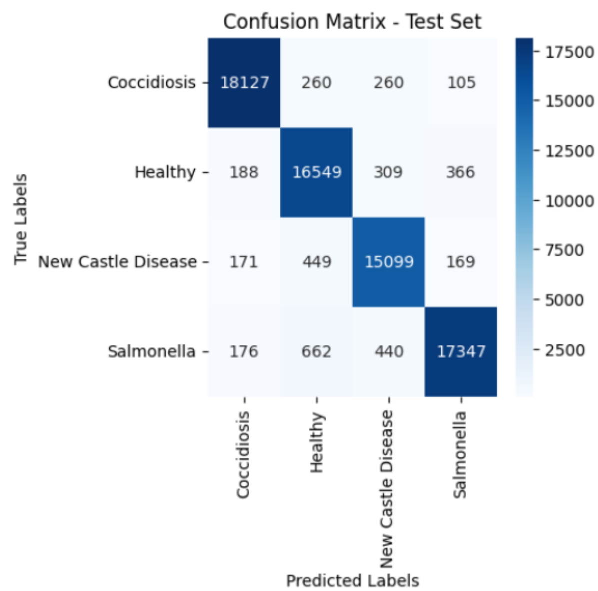

Fig. 9. Confusion matrix of the ResNet18 model on the test set.

VII. Discussion on Model Performance

The performance comparison between ResNet-18, ResNet-34, and ResNet-50 reveals that a deeper architecture does not necessarily guarantee superior results. Contrary to common expectations, the lightweight ResNet-18 outperformed both ResNet-34 and ResNet-50 in terms of accuracy (94.17%) and matched them in F1-score (0.95), while also requiring significantly less training time and computational resources.

This outcome highlights that, in this specific task of classifying poultry feces images, the added depth and complexity of ResNet-50 did not translate into better performance. In fact, ResNet-50 exhibited the lowest accuracy (91.5%) despite being the most computationally intensive model. These findings suggest that the discriminative features present in the dataset—such as texture and color variations—are effectively captured by shallower architectures without the need for excessive depth.

Therefore, ResNet-18 offers the best trade-off between accuracy, training efficiency, and model size, making it the most suitable choice for practical deployment, especially in resource-constrained environments like agricultural settings.

VIII. Conclusion and Future Work

In this study, we have demonstrated the effectiveness of convolutional neural networks, particularly ResNet architectures, in classifying poultry diseases based on fecal images. Among the tested models, ResNet-18 stood out as the best-performing architecture, achieving the highest accuracy (94.17%) while maintaining a favorable balance between model complexity and generalization capability. Despite being shallower than ResNet-34 and ResNet-50, ResNet-18 proved to be more efficient and better suited for the task, especially given the moderately sized dataset and the nature of the visual cues.

Unlike medical imaging domains where structural patterns (e.g., shapes or anatomical landmarks) are critical, fecal image classification relies more heavily on **color variations, textures, and localized abnormalities**, which ResNet-18 was able to capture effectively without requiring excessive model depth. This emphasizes that deeper models are not always superior, particularly when the target features are relatively low-level and the dataset size is limited.

The application of Grad-CAM further enhanced interpretability by visually confirming that the model focuses on biologically relevant regions, such as lesions, discoloration, and abnormal textures, rather than irrelevant background. This adds a layer of transparency to the system and increases its trustworthiness for practical use in poultry farming environments.

For future work, we aim to incorporate additional contextual data—such as environmental conditions, feed composition, or temporal disease progression—to enrich the feature space and improve robustness. Ultimately, this research contributes to the development of accessible, interpretable, and computationally efficient AI-driven tools that support early diagnosis and intervention in poultry health management.

Acknowledgments

I would like to express my sincere gratitude to Chandrashekar Natesh and his team for providing the dataset used in this study [7]. Their contribution was fundamental to the successful implementation and evaluation of the classification models.

I am especially thankful to Professor Ricardo Espinoza for his constant support, valuable guidance, and encouragement throughout the entire duration of this project. His insights into model training and optimization were crucial in overcoming key challenges and achieving meaningful results.

References

[1] K. Srivastava and P. Pandey, "Deep Learning Based Classification of Poultry Disease," *Applied Science and Advanced Materials Technologies*, vol. 13, no. 1, 2023. [Online]. Available: https://gigvvy.com/journals/ausmt/articles/ausmt-2023-13-01-2439

[2] M. Z. Degu and G. L. Simegn, "Smartphone based detection and classification of poultry diseases from chicken fecal images using deep learning techniques," *Procedia Computer Science*, vol. 167, pp. 345–354, 2020. https://www.sciencedirect.com/science/article/pii/S2772375523000515

[3] A. Nakrosis, A. Paulauskaite-Taraseviciene, V. Raudonis, I. Narusis, V. Gruzauskas, R. Gruzauskas, and I. Lagzdinyte-Budnike, "Towards Early Poultry Health Prediction through Non-Invasive and Computer Vision-Based Dropping Classification," *Animals*, vol. 13, no. 19, 2023. [Online]. Available: https://www.mdpi.com/2076-2615/13/19/3041

[4] J.-R. Bisailon, T. E. Feltmate, S. Sheffield, R. Julian, E. Todd, C. Poppe, and S. Quessy, "Classification of Grossly Detectable Abnormalities and Conditions Seen at Postmortem in Canadian Poultry Abattoirs According to a Hazard Identification Decision Tree," *Preventive Veterinary Medicine*, vol. 213, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0362028X22007608

[5] M. S. Uwadia Osagie and F. Cyril-Musa, "Classification of Poultry Birds Based on Health State Using Convolutional Neural Network," *African Journal of Science, Technology, Innovation and Development*, vol. 15, no. 1, 2023. [Online]. Available: https://www.ajol.info/index.php/aujst/article/view/279446

[6] H. Hemalatha, S. Muruganand, and R. Maheswaran, "Recognition of Poultry Disease in Real Time using Extreme Learning Machine," *ICIDRET 2014*, 2014. [Online]. Available: https://edlib.net/2014/icidret/icidret2014008.pdf

[7] C. Natesh, "Poultry Diseases: Visual Dataset for Classification," Kaggle, 2022. [Online]. Available: https://www.kaggle.com/datasets/chandrashekarnatesh/poultry-diseases