InceptionV4-Based Bird Classification For Rice Farm Avian Threat Mitigation Using Drone-Based Base-Station Processing

Itoro Kingsley Akpabio[1]
Department of Electrical/Electronic Engineering
University of Uyo , Akwa Ibom State, Nigeria
Emai: itoroakpabio@uniuyo.edu.ng

Eke Godwin Kelechi[2]
Computer Engineering Department,
Federal Polytechnic Nekede Owerri, FPNO
Imo State, Nigeria.
Email: eke.kelechy@gmail.com

Akpasam Joseph Ekanem[3]
Department of Electrical and Electronic Engineering,
Akwa Ibom State University, Mkpat Enin, Akwa Ibom State
akpasamekanem@aksu.edu.ng; ajekanem56@yahoo.com;
ORCID: 0000-0001-8360-4165

Abstract

This paper presents an InceptionV4-based bird classification method for rice farm avian threat mitigation using drone-based base-station processing. The study focuses on a robust computer vision model deployed at the base station, where higher computational capacity supports the use of a deeper classifier than is practical on lightweight drone hardware. A combined CUB-200-2011 and NABirds dataset subset was prepared for six coded bird classes using an 80:20 training-validation split. The trained InceptionV4 model achieved strong validation performance. Training loss decreased from 2.8 to 0.3, while validation loss decreased from about 2.9 to 0.3 over 40 epochs. The average precision, recall, and F1-score across the six classes were 99.35%, 99.47%, and 99.41%, respectively. Class_5 produced the highest recall of 99.7% and the highest F1-score of 99.65%. The results show that an InceptionV4-based base-station classifier can support accurate bird recognition for intelligent, non-lethal avian threat mitigation in rice farms. Field testing with real rice-farm drone footage is still required to confirm generalization under operational conditions.

Keywords: InceptionV4 model, bird classification, avian threat mitigation, rice farm, drone imagery, base-station processing, computer vision.

1. Introduction

Rice farms are frequently exposed to bird attacks, especially during grain-filling and ripening periods. These attacks reduce yield, increase the cost of manual guarding, and create uncertainty for farmers. Traditional deterrent methods, such as scarecrows, manual chasing, reflective tapes, and noise-making devices, are often labour-intensive and inconsistent. Recent work also shows that deep-learning-based bird detection can be integrated with automated repellent mechanisms for agricultural protection [10]. Traditional deterrent methods also lack the ability to distinguish harmful bird activity from harmless background motion.

Computer vision provides a stronger foundation for targeted avian threat mitigation because it enables automatic recognition of bird categories from images or video frames. When vision models are combined with drone-based monitoring, farms can be observed from wider viewpoints, and detected bird activities can be linked to timely response decisions [9], [10]. However, drone imagery also introduces challenges such as motion blur, changing altitude, small object size, background clutter, and unusual viewing angles.

A key design choice in such a system is the location of the classification model. Onboard drone models are usually constrained by battery capacity, memory, processor speed, and heat dissipation. A base-station model has fewer computational restrictions and can therefore use a deeper architecture to achieve higher recognition accuracy. This makes the base station suitable for detailed bird classification, while the drone can focus on image acquisition, frame selection, and communication.

Fine-grained bird classification is challenging because many bird classes share similar body shapes, colours, feather textures, and poses. InceptionV4 is suitable for this task because its multi-scale convolutional design captures both local and contextual features. This study uses InceptionV4 as the main base-station classifier and retains the reported experimental results for six bird categories. A Capsule Network branch is described as a complementary architectural extension for future viewpoint-robust ensemble implementation, but the reported implementation results in this paper are for the InceptionV4 base-station classifier.

The main contributions of the paper are as follows. First, a drone-to-base-station bird classification workflow is presented for rice farm avian threat mitigation. Second, the dataset preparation and experimental configuration are clarified to improve reproducibility. Third, the InceptionV4 base-station model architecture is described with supporting equations and a camera-ready diagram. Fourth, the retained loss, precision, recall, F1-score, and misclassification results are harmonized with the methodology and discussed using cautious deployment-oriented language.

2. Materials and Methods

2.1 System Model for Base-Station Bird Classification

The proposed system consists of four functional layers: aerial image acquisition, frame and crop preparation, base-station model inference, and avian threat response. A drone captures aerial footage over the rice farm. The video stream is converted into frames, and bird regions are extracted as image crops. These image crops are transferred to the base station, where the InceptionV4 classifier assigns each crop to one of six bird classes. The classification result can then support alert generation, deterrent control, or later retraining of lightweight edge models.

The system is designed as an accuracy-first pipeline. Drone-side processing may perform frame selection or coarse object localization, while the base station performs detailed bird classification using a deeper model. This arrangement is useful because the base station can provide more stable processing power, larger memory, and easier model updating than the onboard platform. The classification output is intended to support non-lethal deterrent actions, such as acoustic, visual, or movement-based responses, rather than harmful treatment of birds.

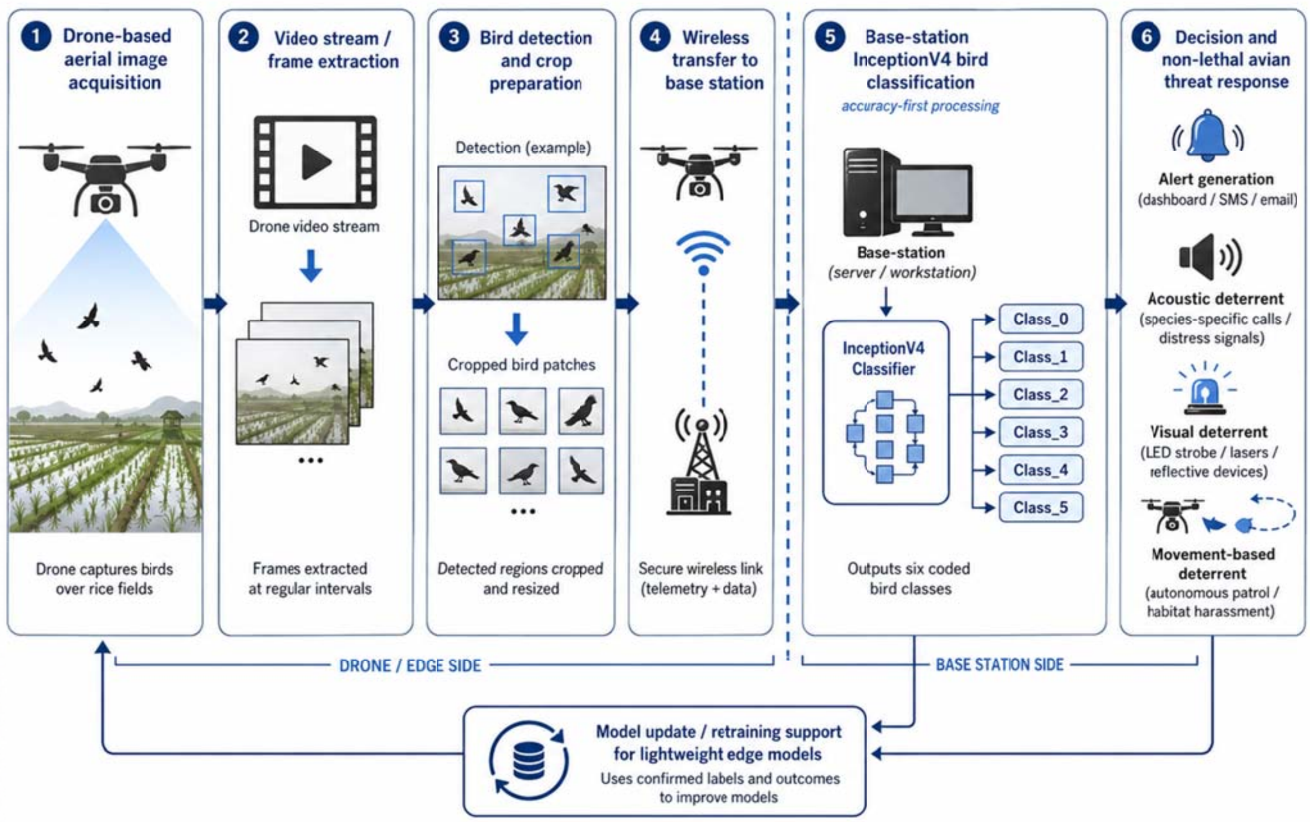

Figure 1. System model architecture for base-station bird classification.

2.2 Dataset Preparation and Preprocessing

The study used a combined bird-image dataset derived from CUB-200-2011 and NABirds. CUB-200-2011 contains 11,788 images across 200 bird categories, while NABirds contains 48,562 images across 555 bird categories [2], [3]. These datasets are useful for fine-grained bird recognition because they provide labelled bird images with visible appearance differences such as shape, colour, feather pattern, and pose. For this study, the combined dataset was reorganized into six experimental bird categories labelled Class_0 to Class_5. The species names behind the six class codes were not available in the retained experimental records; therefore, the coded class names are retained to avoid unsupported relabelling.

The class labels were retained as Class_0 to Class_5 because the original experimental record did not include verified species names. This avoids unsupported relabelling while preserving consistency with the reported results. Future experimental records should map each coded class to its verified bird species or threat category. Images were validated, resized, normalized, and augmented before splitting into training and validation subsets.

Table 1. Six coded bird classes used in the experimental results.

Class code	Class identity	Dataset source
Class_0	Coded avian class 0	CUB-200-2011/NABirds subset
Class_1	Coded avian class 1	CUB-200-2011/NABirds subset
Class_2	Coded avian class 2	CUB-200-2011/NABirds subset
Class_3	Coded avian class 3	CUB-200-2011/NABirds subset
Class_4	Coded avian class 4	CUB-200-2011/NABirds subset
Class_5	Coded avian class 5	CUB-200-2011/NABirds subset

Table 2. Dataset composition used for the InceptionV4 base-station classifier.

Item	Description	Reported value	Remark
Source datasets	CUB-200-2011 and NABirds	Combined bird-image source	Used for fine-grained bird

			recognition
Experimental classes	Coded bird categories	Class_0 to Class_5	Class names retained as reported in the results
Data split	Training and validation partition	80:20	Used for model development and validation
Validation support	Samples per class in the reported result table	333	Balanced validation support
Total validation samples	333 samples x 6 classes	1,998	Derived from the retained results

2.3 Base-Station Classification Framework

The base-station classification pipeline prioritizes accuracy across the overall avian mitigation system. It preprocesses drone frames, classifies bird crops, and produces predictions that can guide deterrent actions or support periodic updating of lightweight edge models. The base-station setup is less constrained by power and memory than onboard processing, making it suitable for a deeper architecture such as InceptionV4.

A Capsule Network branch is shown only as a complementary extension because capsule representations can encode pose-related information and part-whole relationships. This may be useful for future drone-view bird recognition under unusual viewing angles or partial occlusion. The numerical results in this paper remain InceptionV4 results only.

2.4 InceptionV4 Model Architecture

InceptionV4 extends earlier Inception models by using a deeper and more streamlined multi-branch convolutional design [1]. The model combines stem layers, Inception blocks, reduction layers, global average pooling, dropout, and a final dense softmax classifier. The multi-branch design allows the network to learn features at different receptive-field scales. This is useful for bird classification because useful visual cues may appear in small regions, such as beak shape and feather texture, as well as in larger regions, such as body outline and wing shape.

For an input bird crop x, the InceptionV4 feature extractor produces a deep representation h_IV4. After global average pooling, the classifier produces the pooled feature vector z, six-class logits, and class probabilities as follows:

$$\mathbf{h}_{IV4} = f_{IV4}(\mathbf{x}; \boldsymbol{\theta}_{IV4}), \quad \mathbf{z} = GAP(\mathbf{h}_{IV4}), \quad \mathbf{p}_{IV4} = \mathrm{softmax}(\mathbf{W}_{IV4}\mathbf{z} + \mathbf{b}_{IV4}) \tag{1}$$

In (1), x is the input bird crop, theta_IV4 denotes the learned InceptionV4 parameters, h_IV4 is the extracted deep representation, z is the pooled feature vector, W_IV4 represents the learned classification weights, b_IV4 is the learned bias vector, and p_IV4 is the six-class probability vector. The multi-scale Inception block concatenates features extracted by different branches as follows:

$$\mathbf{H}_{out} = \mathrm{Concat}\big(\mathbf{B}_1(\mathbf{x}), \mathbf{B}_2(\mathbf{x}), \dots, \mathbf{B}_n(\mathbf{x})\big) \tag{2}$$

This concatenation increases feature diversity without excessive parameter growth. Batch normalization is applied to stabilize training as follows [5]:

$$\hat{x}^{(k)} = \frac{x^{(k)} - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}, \quad y^{(k)} = \gamma \hat{x}^{(k)} + \beta \tag{3}$$

Label smoothing is used to reduce overconfidence in the fine-grained classes as follows [6]:

$$y_k^{LS} = (1 - \alpha) y_k + \frac{\alpha}{K} \tag{4}$$

In (4), K is the number of classes, alpha is the label smoothing factor, y_k is the original one-hot target for class k, and y_k^LS is the smoothed target value.

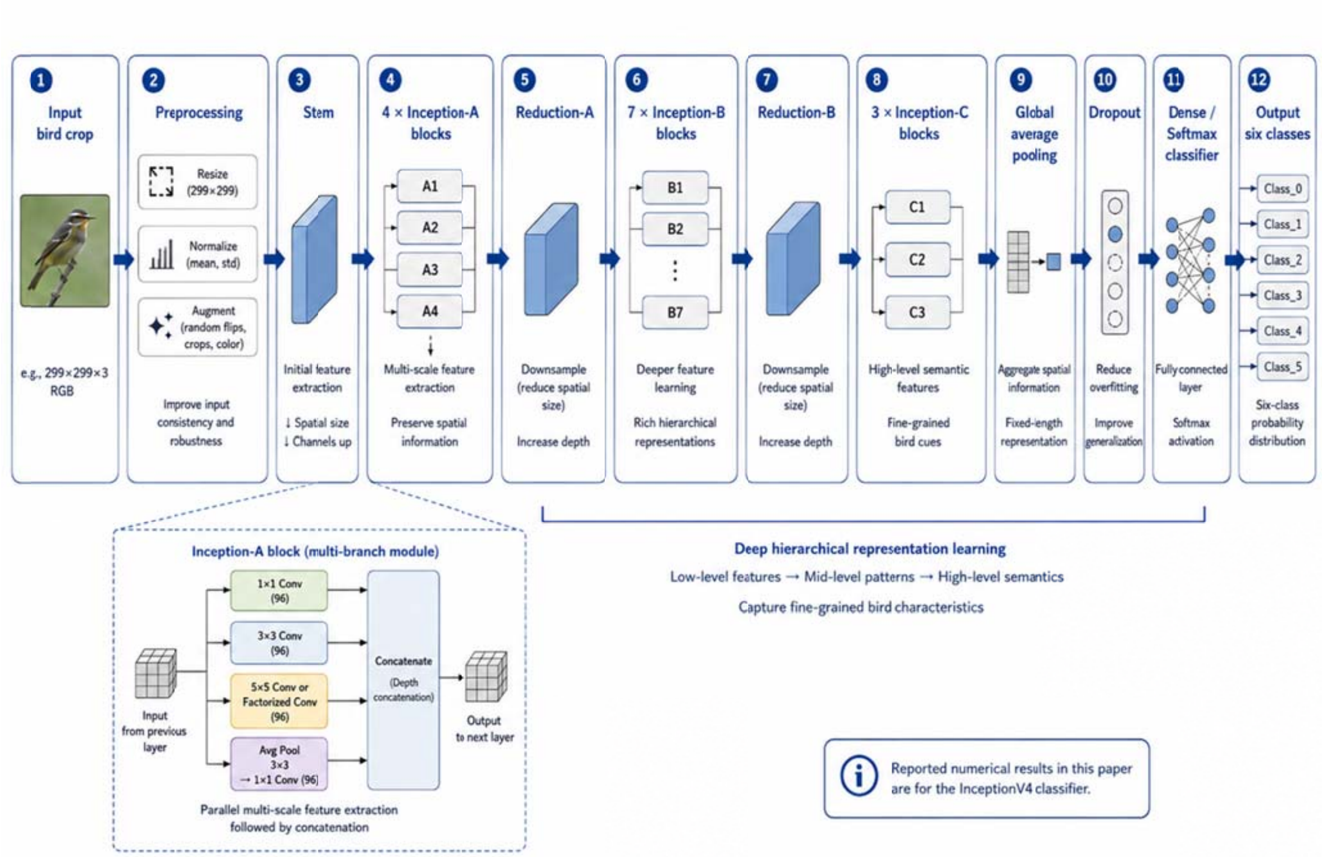

Figure 2. InceptionV4 model architecture used for fine-grained bird classification.

2.5 Capsule Network Extension

A Capsule Network branch is retained as a future viewpoint-robust extension of the base-station design [4]. Its role is to represent pose-related information and part-whole relationships that may occur when drone images capture birds from different angles or under partial occlusion.

For completeness, the capsule extension can be expressed through prediction vectors, routing coefficients, capsule inputs, squashing, and final class selection as follows:

$$\hat{\mathbf{u}}_{j|i} = \mathbf{W}_{ij}\mathbf{u}_i \quad (5)$$

$$c_{ij} = \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})} \quad (6)$$

$$\mathbf{s}_j = \sum_i c_{ij}\, \hat{\mathbf{u}}_{j|i} \quad (7)$$

$$\mathbf{v}_j = \frac{\|\mathbf{s}_j\|^2}{1+\|\mathbf{s}_j\|^2} \frac{\mathbf{s}_j}{\|\mathbf{s}_j\|} \quad (8)$$

$$\hat{y} = \underset{j}{\operatorname{argmax}} \parallel \mathbf{v}_j \parallel \quad (9)$$

In (5)-(9), u_i is the lower-level capsule output, W_ij is the transformation matrix, c_ij is the routing coefficient, s_j is the total input to capsule j, v_j is the squashed output vector, and ŷ is the selected class label.

The reported numerical results are interpreted strictly as InceptionV4 results. CapsNet is not evaluated separately in the retained result set; it is included only to show a possible future extension for stronger viewpoint handling.

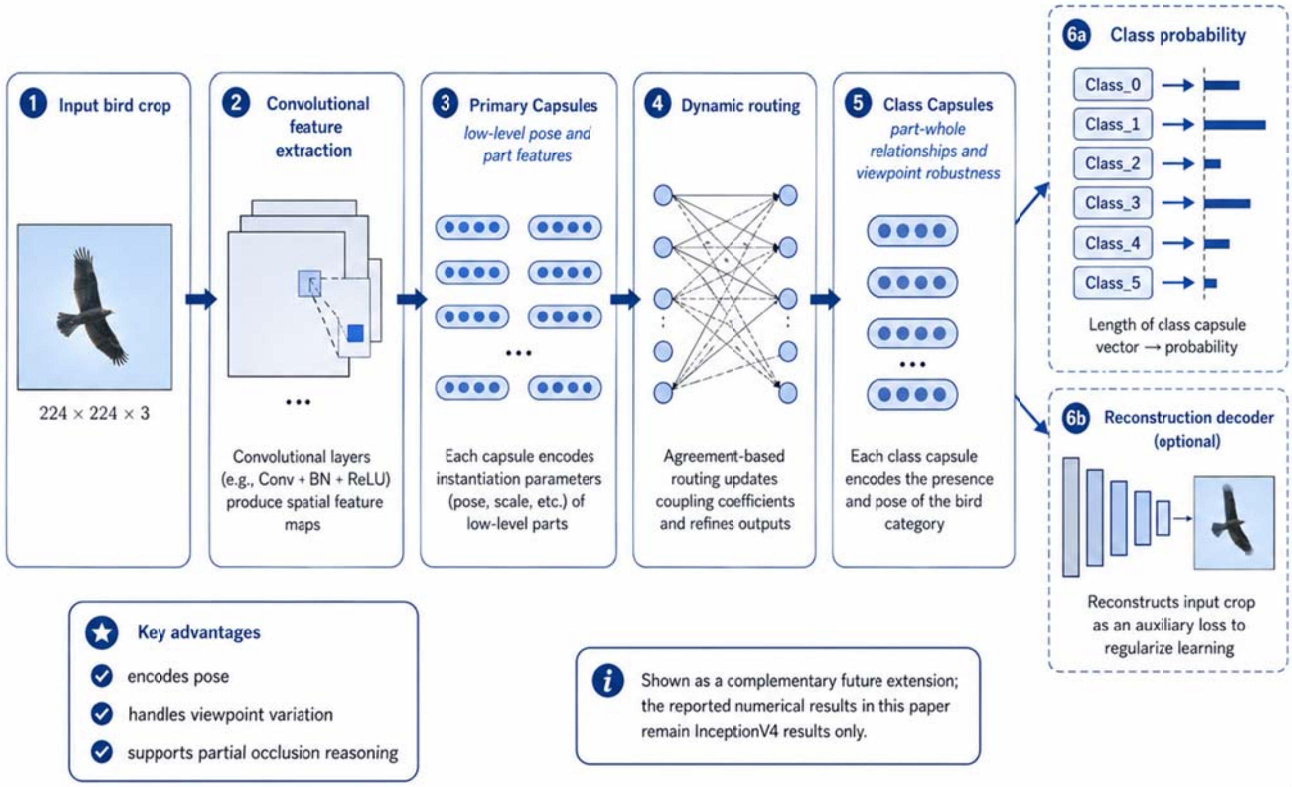

Figure 3. Capsule Network architecture for viewpoint-robust bird classification.

2.6 Training and Evaluation Procedure

The InceptionV4 model was initialized with pretrained weights and fine-tuned on the prepared six-class bird crop dataset. The final classification layer was replaced with a six-class softmax head. Training used the prepared 80% training subset, while model selection and reporting used the 20% validation subset. Loss values were recorded for 40 epochs. Precision, recall, F1-score, and notable misclassification counts were then computed from the validation predictions.

Table 3 summarizes the experimental configuration for the base-station InceptionV4 classifier. It separates the retained experimental details from implementation settings used to support reproducible fine-tuning and validation.

Table 3. Experimental configuration for the base-station InceptionV4 classifier.

Configuration item	Value/setting	Purpose
Main classifier	InceptionV4	Base-station six-class bird classification
Input data	Prepared bird image crops	Classification of detected bird regions
Number of classes	6	Class_0 to Class_5
Data split	80% training and 20% validation	Model training and validation
Validation support	333 samples per class	Balanced per-class reporting
Total validation samples	1,998	Six classes x 333 samples
Training epochs	40	Loss curve reporting
Loss function	Cross-entropy with optional label smoothing	Multi-class classification
Optimizer	Adam	Stable fine-tuning of pretrained weights
Initial learning rate	0.0001	Conservative transfer-learning update
Batch size	32	Mini-batch training
Input image size	299 x 299 x 3	Standard InceptionV4 image input
Data augmentation	Horizontal flip, rotation, zoom, brightness adjustment and rescaling	Robustness to drone-view variations
Hardware/software environment	GPU-enabled Python deep-learning environment	Base-station model training and evaluation
Evaluation metrics	Precision, recall, F1-score and misclassification count	Performance assessment

The classification workflow followed the algorithm below.

Algorithm 1. Bird Classification Pipeline for the Base-Station Model

Input: prepared crop dataset D, pretrained InceptionV4 weights θ_IV4, six output classes, learning parameters, and validation set.

Output: trained InceptionV4 base-station classifier and predicted bird class ŷ.

1. Load and validate the prepared bird crop images.
2. Remove unusable images and encode labels from Class_0 to Class_5.
3. Resize, normalize, and augment the training images.
4. Split the dataset into 80% training and 20% validation subsets.
5. Initialize InceptionV4 with pretrained weights.
6. Replace the final layer with a six-class softmax classification head.
7. Train the model over 40 epochs using mini-batches from the training set.
8. Record training and validation losses at each epoch.
9. Compute validation precision, recall, F1-score, and notable misclassification cases.
10. Select the trained model for base-station bird classification.

2.7 Evaluation Metrics

The evaluation used precision, recall, F1-score, and misclassification count. Precision measures the proportion of predicted samples in a class that truly belong to that class. Recall measures the proportion of actual samples in a class that are correctly identified. F1-score gives the harmonic mean of precision and recall.

$$\text{Precision}_c = \frac{TP_c}{TP_c + FP_c} \quad (10)$$

$$\text{Recall}_c = \frac{TP_c}{TP_c + FN_c} \quad (11)$$

$$F1_c = \frac{2\,\text{Precision}_c\,\text{Recall}_c}{\text{Precision}_c + \text{Recall}_c} \quad (12)$$

3. Results and Discussion

3.1 Training and Validation Loss Behaviour

The results for the implementation and evaluation of the base-station InceptionV4 bird classification model are presented in Figure 4, Table 4, and Table 5. The training and validation loss curves show steady improvement over 40 epochs. The training loss decreased from 2.8 to 0.3, while the validation loss decreased from about 2.9 to 0.3. The close alignment between the two curves suggests stable learning and no strong evidence of overfitting under the reported validation setting.

The loss pattern is important because fine-grained bird classes can be difficult to separate when the images contain similar colours, poses, or textures. A large gap between training and validation losses would suggest that the model memorized the training data. In this case, the retained results show that both curves decreased together, which supports the use of the model as an accuracy-focused base-station classifier.

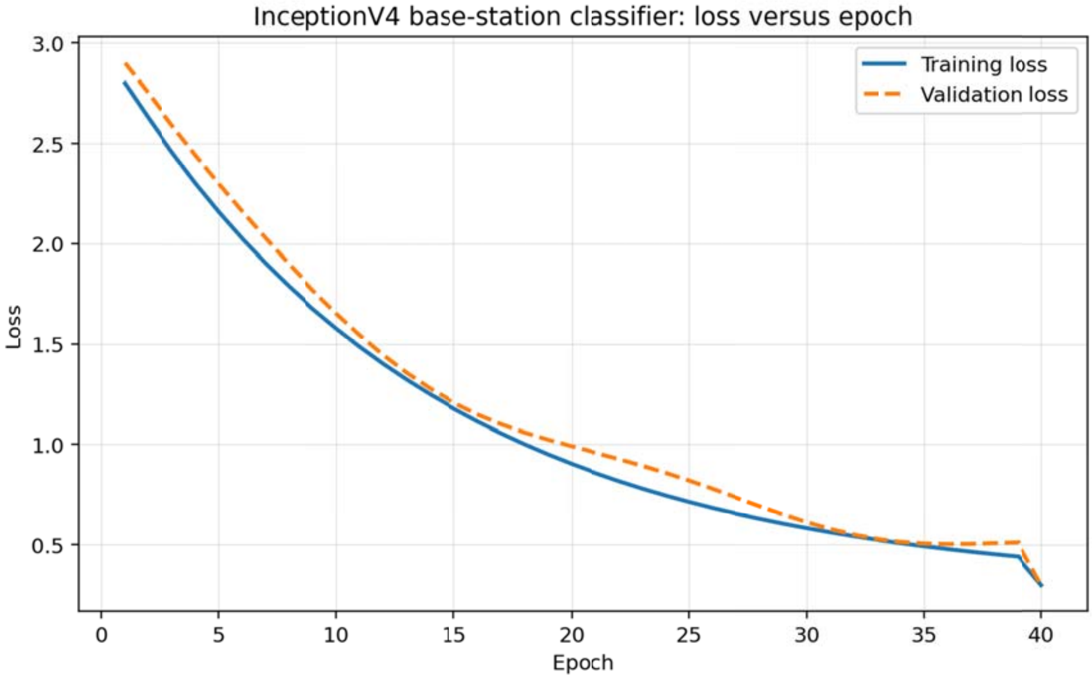

Figure 4. Training and validation loss curves of the InceptionV4 base-station bird classification model over 40 epochs.

3.2 Per-Class Classification Performance

The per-class results show consistently high performance across all six bird categories. The highest precision values were obtained for Class_2 and Class_5, with both classes achieving 99.6%. Class_5 also achieved the highest recall of 99.7% and the highest F1-score of 99.65%. Class_4 had the lowest precision at 98.9% and the lowest F1-score at 99.00%, although this value is still high for a six-class fine-grained classification task.

Table 4. Per-class result for base-station classification model using InceptionV4.

Class	Support	Precision (%)	Recall (%)	F1-score (%)
Class_0	333	99.4	99.7	99.55
Class_1	333	99.1	99.3	99.20
Class_2	333	99.6	99.6	99.60
Class_3	333	99.5	99.4	99.45
Class_4	333	98.9	99.1	99.00
Class_5	333	99.6	99.7	99.65
Average	**333.00**	**99.35**	**99.47**	**99.41**

The average precision, recall, and F1-score were 99.35%, 99.47%, and 99.41%, respectively. These results indicate that the base-station InceptionV4 classifier achieved balanced recognition performance across the six coded classes. The high average precision means that few predicted class assignments were wrong, while the high average recall means that most actual class instances were correctly detected within the validation set.

3.3 Misclassification Analysis

The notable misclassification cases are shown in Table 5. The highest listed error occurred when Class_4 was predicted as Class_1, with a count of 2. Another case occurred when Class_1 was predicted as Class_4, with a count of 1. These errors suggest that Class_1 and Class_4 have some overlapping visual features in the validation samples. This may be due to similar body shape, pose, background appearance, or feather pattern under the prepared image conditions.

Table 5. Notable misclassification cases observed for the InceptionV4 base-station model.

True class	Predicted class	Count
Class_4	Class_1	2
Class_1	Class_4	1

Table 5 reports only the notable error pairs available in the retained result summary. A full confusion matrix was not available in the retained result sheet; therefore, a complete class-to-class error distribution should be reported in future field validation work when all prediction records are available.

3.4 Discussion of Deployment Implications

The retained results support the use of InceptionV4 as a strong base-station classifier for bird recognition. In practical farm deployment, a wrong class assignment may trigger unnecessary deterrent actions or may fail to prioritize a high-risk bird category. Deterrent action should therefore be triggered only when the predicted class confidence exceeds a predefined threshold determined during field calibration. This threshold is important for reducing false alarms when non-threatening objects, distant birds, or unclear image crops are processed by the system. The high validation precision and recall provide a useful starting point for intelligent, non-lethal avian threat mitigation. The base station can also serve as a trusted model-update point for lighter onboard models.

The result should still be interpreted with care because real farms may introduce weather variation, low light, rain, motion blur, occlusion by rice plants, mixed bird groups, and different camera heights. Field testing with real rice-farm drone footage is recommended to confirm robustness under practical farm conditions.

4. Conclusion

This paper presented an InceptionV4-based bird classification method for rice farm avian threat mitigation using drone-based base-station processing. The system model shows how drone-acquired images are processed at the base station for detailed six-class bird recognition. The classification output is intended to support non-lethal deterrent responses such as acoustic, visual, or movement-based actions.

The reported experimental results show that the base-station InceptionV4 model achieved strong classification performance. Training loss decreased from 2.8 to 0.3, while validation loss decreased from 2.9 to 0.3 over 40 epochs. The average precision, recall, and F1-score were 99.35%, 99.47%, and 99.41%, respectively. Class_5 produced the best F1-score of 99.65%, while the notable misclassification cases mainly involved Class_1 and Class_4.

These outcomes indicate that an InceptionV4-based base-station classifier can support accurate bird recognition for intelligent avian threat mitigation in rice farms. The claim is limited to the reported validation setting. Further field testing with real rice-farm drone footage is recommended to confirm robustness under practical farm conditions. Future work should also include verified mapping of Class_0 to Class_5 to bird species or threat categories, a full confusion matrix, complete optimizer and hardware reporting, and direct comparison with lightweight onboard models.

References

[1] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning," in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, 2017, pp. 4278-4284, doi: 10.1609/aaai.v31i1.11231.

[2] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie, "The Caltech-UCSD Birds-200-2011 Dataset," California Institute of Technology, Tech. Rep. CNS-TR-2011-001, 2011.

[3] G. Van Horn et al., "Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 595-604, doi: 10.1109/CVPR.2015.7298658.

[4] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in Advances in Neural Information Processing Systems, vol. 30, 2017, pp. 3856-3866.

[5] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in Proceedings of the 32nd International Conference on Machine Learning, 2015, pp. 448-456.

[6] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception architecture for computer vision," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818-2826, doi: 10.1109/CVPR.2016.308.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017, doi: 10.1145/3065386.

[8] T. Lin et al., "Focal loss for dense object detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 2, pp. 318-327, 2020, doi: 10.1109/TPAMI.2018.2858826.

[9] A. Coluccia et al., "Drone vs. bird detection: Deep learning algorithms and results from a grand challenge," Sensors, vol. 21, no. 8, Art. no. 2824, 2021, doi: 10.3390/s21082824.

[10] Y.-C. Chen, J.-F. Chu, K.-W. Hsieh, T.-H. Lin, P.-Z. Chang, and Y.-C. Tsai, "Automatic wild bird repellent system that is based on deep-learning-based wild bird detection and integrated with a laser rotation mechanism," Scientific Reports, vol. 14, Art. no. 15924, 2024, doi: 10.1038/s41598-024-66920-2.

[11] M. K. Baowaly et al., "Deep transfer learning-based bird species classification using mel spectrogram images," PLoS ONE, vol. 19, no. 8, Art. no. e0305708, 2024, doi: 10.1371/journal.pone.0305708.

[12] P. Marcoň et al., "A system using artificial intelligence to detect and scare bird flocks in the protection of ripening fruit," Sensors, vol. 21, no. 12, Art. no. 4244, 2021, doi: 10.3390/s21124244.