BanglaBirds-AttnNet: A Framework for Classification Endangered Bangladeshi Birds Using EfficientNetB0 with CBAM Enhanced By Explainable AI

Md. Abu Raihan[1] and Sadia Bristi[1]

[1] Department of Computer Science and Engineering, Khwaja Yunus Ali University, Enayetpur, Chauhali, Bangladesh
raihan.cse@kyau.edu.bd*, sadiabristi031@gmail.com

Abstract. Birds are vital indicators of ecosystem health, yet numerous species in Bangladesh are threatened by habitat loss and environmental change. Existing bird classification models often employ black-box architectures lacking interpretability and dataset specificity. This paper introduces BanglaBirds-AttnNet, a deep learning framework that combines EfficientNetB0 with a Convolutional Block Attention Module (CBAM) to enhance spatial-channel feature learning for bird classification. Trained on the BanglaBirds dataset, which contains 18 native and endangered species, the proposed model achieved 99% classification accuracy, outperforming MobileNet, ViT, and DarkNet- based approaches. The inclusion of explainable AI improves the transparency and interpretability of predictions, enabling reliable real-world deployment for ecological monitoring. BanglaBirds-AttnNet thus represents a significant advancement in AI-driven biodiversity conservation, delivering both high accuracy and explainability for endangered bird classification in Bangladesh.

Keywords: BangladeshiBirds, Endangered Species Classification, EfficientNetB0, Convolutional Block Attention Module (CBAM), Deep Learning, Explainable Artificial Intelligence (XAI)

1 Introduction

Birds are among the most important bioindicators of environmental health and stability of ecosystems. But in Bangladesh,both native and migratory bird species are increasingly under threat amid rapid habitat loss, urbanisation and climate change. The demand of automated bird identification systems have increased substantially to support biodiversity monitoring programs and conservation measures. Recent developments in deep learning have achieved striking success for fine-grained bird classification, most existing models are trained on global datasets (e.g., CUB-200 and NABirds), which are not well representative of the endangered species native to Bangladesh. Previously, MobileNet and

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_53

transfer learning techniques have been shown to provide efficient lightweight classification[9], while Vision Transformers (ViT) are demonstrated to give strong feature extraction[8]. Hybrid architectures which combine CNN with Transformer have as well achieved competitive accuracy[7]. However, three key limitations persist. In the first place, there is no regional datasets developed for native species. Second, there is not enough employment of attention components to enfold subtle morphological disparity across species. Lastly, the incomprehensibility of model prediction is a serious problem urgently requiring attention and effort in terms of both ecological validation and policy-based conservation. For overcoming these issues, in this paper,we propose a new interpretable deep learning framework, **BanglaBirds-AttnNet,** that incorporates **EfficientNetB0** and the recently proposed **Convolutional Block Attention Module (CBAM)**. The introduced approach makes use of channel and spatial attention together for improved discriminative feature learning; meanwhile, it is computationally efficient for deployment in the field. Our key contributions are summarized as follows:

(i) We augment the publicly available BanglaBirds[11] dataset with additional data and creating a more localized resource for studying native and endangered bird species of Bangladesh.
(ii) We introduce an effective combination of EfficientNetB0 and CBAM for better feature representation learning and attention map generation to capture subtle morphological differences.
(iii) Our model adds interpretability through visual attention maps, which are essential for ecological validation and policy-guided conservation.

The rest of this paper is organized as follows. Section 2 reviews related work on bird classification and attention mechanisms. Section 3 describes the methodology, focusing on the BanglaBirds-AttnNet framework. Section 4 covers the experimental results and evaluation. Section 5 provides a discussion of the findings. Section 6 concludes with key insights and future research directions.

2 Related Work

Prior studies on Bangladeshi bird classification have almost exclusively focused on small subsets of species and have been limited by the capacities of their datasets, resulting in rather mediocre performances. The majority of previous works have used traditional CNN-based structures; however, the use of state-of-the-art architectures (such as EfficientNet) and attention mechanisms, including CBAM, has not been studied. To fill this gap, we present BanglaBirds-AttnNet that takes the advantage of EfficientNetB0 architecture with CBAM attention to facilitate feature representation and produce better classification performance for endangered Bangladeshi birds.

Lin et al.[4] initially proposed Bilinear CNN models for the fine-grained visual recognition task, which extends single convolutional layer pairs with outer product-based irradiance features and a pooling layer to capture cascaded local pairwise feature interactions. After that, the pre-trained model was adapted to the CUB-200-2011 dataset with a classification accuracy of 84.1%. The approach demonstrated simplicity in training and competitiveness, as well as fast inference speed (8 fps on a Tesla K40 GPU). However, it was evaluated only on a single dataset. Their method also incurred hardware-dependent performance, which makes it difficult to generalize beyond a particular architectural setting.

Rahman et al.[9] proposed a local bird recognition system for Bangladesh using MobileNet and Inception-v3 with and without transfer learning. The one with transfer learning of MobileNet+ achieved the highest performance, reaching 91% compared to other methods. Models were compared using accuracy, F1-score, and Receiver Operating Characteristic (ROC) curves, among others, to determine the best model for local bird classification. However, the number of datasets and diversity were unknown, subclasses of birds were restricted, and real-time deployment issues had not been faced.

Ahmed et al.[8] proposed BIRDS-BD, a fine-grained bird classification database of 16 species of Bangladesh with 12,800 images, and used Vision Transformer (ViT) with transfer learning. The proposed approach also achieved substantial precision, recall, and F1-score, with an overall accuracy of 92.03%. This work contributes to wildlife research and conservation by filling in the gap of locally specific data. Nevertheless, the dataset augmentation was not conducted, performance per species was not specified, real-time deployment was not demonstrated, and scaling to more species is unproven.

Woo et al.[15] presented the CBAM (Convolutional Block Attention Module), a lightweight attention mechanism that applies channel-wise as well as spatial-wise attention to CNNs. It achieved superior performance on the ImageNet-1K, MS COCO, and VOC 2007 datasets, and can be easily incorporated into CNN architectures with little computational cost. The latter method, however, is restricted to feed-forward, deeper CNNs, lacks fine-grained precision recall, and its potential for general use beyond this model type and into real-world domains in healthcare has not been explored.

Tan and Le et al.[12] introduced EfficientNet, a model scaling method for CNNs that includes depth, width, and resolution. The EfficientNet-B7 achieves stateof-the-art 84.4% top-1 and 97.1% top-5 accuracy on ImageNet without using adversarial or noisy training set, which demonstrates the effectiveness of our new scaling method. The models also demonstrate excellent transfer learning performance on CIFAR-100 and Flowers. However, no mention is made of training computational costs and hardware requirements, nor are precision/recall metrics reported for tasks beyond ImageNet, which hampers assessment in different applications.

Hoa Le Duc et al.[2] applied EfficientNetB2 with transfer learning for 84 bird species classification, achieving 93% accuracy on both validation and testing sets. The approach improved performance and reduced training time. However, it was limited to 84 species, lacked precision and recall metrics, and its effectiveness may depend on dataset quality, potentially limiting generalization.

Smith et al.[17] used transfer-learning-based EfficientNetB2 for the classification of 84 bird species with a 93% accuracy on both validation and test sets. The method resulted in improved accuracy and reduced training time. You've also found the hierarchical contribution, but it was only evaluated on 84 species, and no precision and recall have been reported, so its performance may depend on the quality of your dataset, which could potentially limit generalization.

Wang et al.[14] introduced a novel fine-grained bird classification with attention-driven data augmentation and discrete knowledge distillation for model compression. The model attained the accuracy of 87.6% footnote.The reported baseline is 88.3%, using only 33% of the parameters, and needed just 1.2G computation for a faster inference with performance preserved at the state-of-the-art level. Nevertheless, there were no precision and recall scores available, and it may not generalize beyond fine-grained bird classification. At the same time, the trade-off between compression rate and interpretability was not examined.

Teterja et al.[13] conducted a performance comparison of lightweight deep learning methods for bird classification, drawing attention to EfficientNetB0, which achieved more than 97% accuracy with a fast response time and low computational cost. The use-case study perceives the suitability of lightweight CNNs in conservation biology and biodiversity monitoring as they can be deployed on edge gateways. However, no precision and recall measurements are provided, the definition of the dataset is unclear, and the model lacks robustness in real-world situations.

In Zhuang et al.[5] a compressive sensing framework was developed for storing and transferring bird image and audio data. The model possessing high reconstruction capabilities, the average PSNR values of the reconstructed images were 33.62 dB (image), 55.76 dB (Mel), and 38.59 dB (WT) under 50% compression, achieving robustness among other conditions, but not limited to, those with less than 30dB under low-rate compressions. Although the model saved the data effectively, it did not return classification metrics (accuracy, precision, or recall). Analysis was limited to 45 bird species with low samples, and the study was mainly emphasized in terms of compression performance, not biological recognition or deployment at a large scale.

For the classification of fine-grained images at the genus level of bats, this study adopted VGG16-CBAM[1] to distinguish seven horseshoe bat taxa residing in Southern China and obtained an accuracy rate of 92.15%. The study proposed a standard photographic protocol and employed Grad-CAM visualizations to demonstrate that the model concentrated on important taxonomic facial regions, making it suggestive of taxonomy automation. Precision and recall were not reported, as the dataset was only 879 images, the images were taken in a controlled setting, and the generalisability to other bat species' wild settings remained unproven.

The bird species recognition research, utilizing a mixed CNN, pioneered an integrated network that combines dense-residual connections, hierarchical structures with depthwise separable convolutions (Inception), efficient channel attention, and an adaptive convolution for spatial weighting. The model was also tested on a 525-bird species dataset and achieved an accuracy of 94%, outperforming Inception V2, ResNet101, DenseNet-264, MobileNetV3-Large, and EfficientNet[16]. Although it showed high accuracy with a low number of parameters and efficiency gains, precision and recall were not shown. The evaluation was confined to a single dataset, and there is no discussion about inference speed or edge deployment, and generalization to real-world bird images has yet to be demonstrated.

3 Methodology

This section discuss about the overall implementation of the study.The total patrs of description given below.

3.1 Dataset Collection and Augmentation

The experiment uses the BanglaBirds [11] dataset, which can be found on Kaggle. The dataset contains 2,700 processed images of 18 bird species in Bangladesh, with 105 images per species. The pictures taken with mobile phones and professional cameras are 224 × 224 pixels in size. At first, the data was divided into training (70%), testing (20%), and validation (10%) datasets, resulting in 33,676 images for training, 10,671 images for testing, and 5,339 images for validation. Then, to improve generalization and reduce overfitting, several on-the-fly augmentation strategies were employed (i.e., random horizontal flipping, rotation up to ±10%, zooming up to ±15%, and contrast adjustment by 10%). This augmentation helps reduce overfitting in the proposed model.

3.2 CBAM: Convolutional Block Attention Module

The **Convolutional Block Attention Module (CBAM)** enhances feature representations by sequentially applying *channel* and *spatial* attention. Given a feature map $F \in R^{C \times H \times W}$, channel attention is computed using both average and max pooling, followed by a shared multi-layer perceptron (MLP) to generate attention weights:

$$M_c(F) = \sigma(MLP(F^c_{avg}) + MLP(F^c_{max})) \tag{1}$$

The refined feature map becomes:

$$F' = M_c(F) \otimes F \tag{2}$$

Spatial attention then highlights salient regions by applying pooling along the channel dimension and a 7 × 7 convolution:

$$M_s(F') = \sigma(f^{7\times7}([AvgPool(F'); MaxPool(F')])) \tag{3}$$

The final output is:

$$F'' = M_s(F') \otimes F' \tag{4}$$

This dual-attention process effectively refines inter-channel dependencies and spatial context, enabling the model to focus on subtle morphological cues essential for fine-grained bird classification and explainable predictions.

3.3 Model Architecture

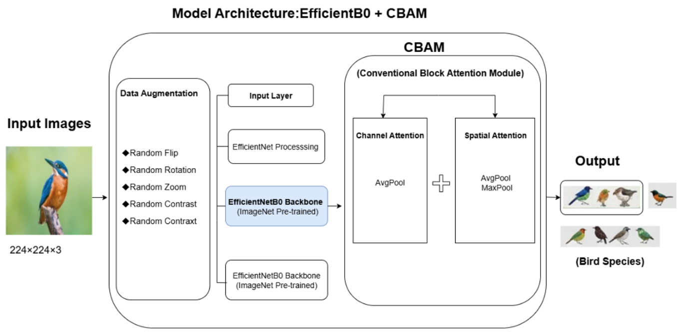

Fig. 1. Proposed BanglaBirds-AttnNet architecture combining EfficientNetB0 with CBAM

The proposed **BanglaBirds-AttnNet** framework figure 1 employs **EfficientNetB0** as a lightweight feature extraction backbone integrated with the **Convolutional Block Attention Module (CBAM)** for adaptive attention refinement. The EfficientNetB0 backbone, pretrained on ImageNet and truncated at the final classification block (*include_ top=False*), extracts a deep hierarchical representation:

$$\mathbf{F}_b = f_{EffNetB0}(\mathbf{X}), \quad \mathbf{F}_b \in R^{H\times W\times C} \tag{5}$$

where H, W, and C denote the spatial height, width, and channel dimensions of the extracted feature map, respectively. The resulting feature tensor $\mathbf{F}_b$ is refined through the CBAM module, which sequentially applies channel and spatial attention mechanisms:

$$\mathbf{F}_{cbam} = \mathcal{A}_s(\mathcal{A}_c(\mathbf{F}_b)) \odot \mathbf{F}_b \tag{6}$$

where $\mathcal{A}_c(\cdot)$ and $\mathcal{A}_s(\cdot)$ denote channel and spatial attention operations, respectively, and $\odot$ represents element-wise multiplication. Subsequently, the attention-refined feature maps are globally aggregated using a Global Average Pooling (GAP) layer:

$$\mathbf{z} = GAP(\mathbf{F}_{cbam}), \quad \mathbf{z} \in R^C \tag{7}$$

A Dropout layer with rate $p = 0.4$ is then applied to mitigate overfitting, followed by a fully connected layer with softmax activation to produce class probabilities:

$$\hat{\mathbf{y}} = softmax(\mathbf{W}\mathbf{z} + \mathbf{b}) \tag{8}$$

where $\mathbf{W}$ and $\mathbf{b}$ denote the learnable parameters of the classifier. This modular design leverages EfficientNetB0's compound scaling efficiency and CBAM's attention refinement to achieve discriminative and interpretable feature learning. The resulting model demonstrates high accuracy with low computational overhead, making it suitable for fine-grained and real-time bird classification tasks in resource-constrained environments.

3.4 Training Strategy

Training was performed with a two-step fine-tuning schedule:
Stage 1 (Warm-up training): During this phase, the entire EfficientNetB0 backbone was kept frozen, allowing only the CBAM and classification layers to train for 12 epochs. The learning rate was set to 1×10^{-3}. This step stabilized training and allowed the model to adapt attention weights specific to the dataset.
Stage 2 (Fine-tuning): In the second phase, the top 50% of EfficientNetB0 layers were unfrozen, while the lower layers remained frozen to retain general visual features. Fine-tuning was performed for 8 additional epochs using a smaller learning rate of 5×10^{-5}. This strategy effectively balanced generalization and domain-specific adaptation.
All experiments were implemented in PyTorch and executed on an NVIDIA GPU environment to ensure efficient computation.

3.5 Explainable AI (XAI) and Visualization Analysis

To enhance interpretability, the proposed **BanglaBirds-AttnNet** integrates Gradient-weighted Class Activation Mapping (Grad-CAM)[10] to visualize class-discriminative regions within bird images. Given a target class c and feature maps A^k, the importance weights are defined as: The class-specific importance weights for each feature map are calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{9}$$

where α_k^c represents the importance of feature map A^k for class c, y^c is the score for class c, and Z is the normalization factor corresponding to the spatial dimensions of the feature map.

The Grad-CAM heatmap for class c is then computed as:

$$L_{Grad-CAM}^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{10}$$

where $ReLU$ ensures that only features with a positive influence on the class of interest are visualized. This heatmap highlights the spatial regions that contribute most to the prediction and is superimposed on the original image for visual interpretation.

The visualization pipeline produces Grad-CAM heatmaps, overlay maps, and probability charts, allowing inspection of attention focus and confidence distribution. Comparative grids illustrate the model's attention consistency across multiple species. Experimental findings reveal that the model predominantly attends to *biologically relevant regions* such as beaks, wings, and plumage rather than background areas.

Thus, **BanglaBirds-AttnNet** not only achieves superior classification accuracy but also provides transparent, explainable decision-making to support biodiversity monitoring and ecological research.

4 Experimental Results and Analysis

Table 1. Test Evaluation Results

Class	**Prec.**	**Rec.**	**F1**	**Supp.**
Asian Koel (Kokil)	0.98	0.81	0.89	573
Black Drongo (Kalo Finge)	0.92	0.93	0.95	582
Black Winged Kite (Chil)	0.99	1.00	0.99	575
Common Myna (Shalik)	0.99	0.99	0.99	582
Common Tailorbird (Tuntuni)	0.99	0.99	0.99	582
Coppersmith Barbet (Basanta Bouri)	0.99	0.99	0.99	584
Heron (Bok)	1.00	0.99	1.00	583
House Crow (Kak)	0.99	0.99	0.92	583
Indian Roller (Neelkanth)	0.99	1.00	0.99	577
Kingfisher (Machranga)	1.00	0.99	1.00	584
Little Cormorant (Pankouri)	0.95	0.99	0.97	584
Oriental Magpie Robin (Doel)	0.99	0.99	0.99	580
Owl (Pecha)	0.99	0.99	0.99	588
Parrot (Tiya)	0.99	0.99	0.99	584
Red Vented Bulbul (Bulbuli)	1.00	0.99	1.00	585
Red Wattled Lapwing (Lal Latika Hottiti)	0.99	1.00	0.99	585
White Breasted Waterhen (Dahuk)	0.99	0.99	0.99	584
White Rumped Shama (Shama)	0.99	0.99	0.99	585
Accuracy	0.99			10477
Macro avg	0.99	0.99	0.99	10477
Weighted avg	0.99	0.99	0.99	10477

The modified CBAM-enhanced EfficientNetB0 was tested on a balanced dataset of 18 Bangladeshi bird species, comprising total 10477 images on testing set. Performance evaluation was conducted with the metrics of overall accuracy, precision, recall, F1-score, and confusion matrix analysis. The model achieved an overall accuracy of 99%. Additionally, the macro-averaged precision, recall, and F1-score exceeded 0.98. These results demonstrate that the architecture is effective in learning inter-class and intra-class variations.

The confusion matrix Figure 2 shows that the model performs well across the 18 bird species. Of the 18 species, in only four cases were misclassifications observed, and no subclasses had an overfitting problem. Misidentifications were mainly observed between similarly looking species like Asian-Koel (Kokil)

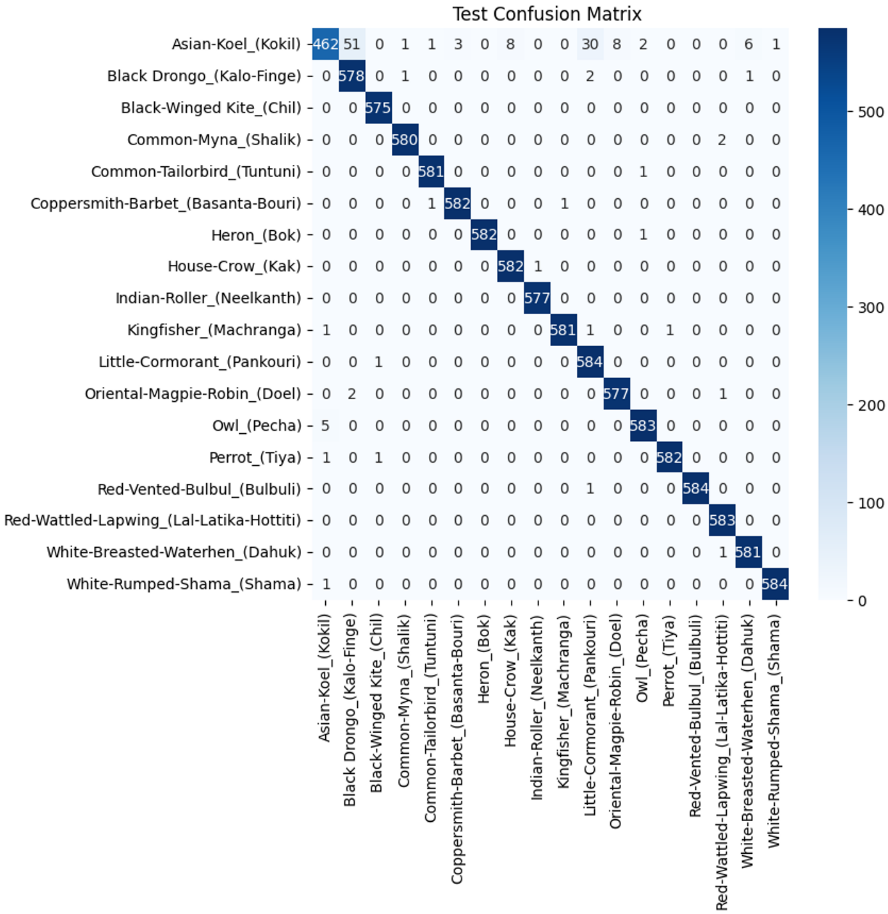

Fig. 2. Confusion Matrix

and Black Drongo (Kalo-Finge). Asian-Koel (Kokil) was misclassified as Black Drongo (Kalo-Finge) 51 times; for example, the error rate is still on the lower side compared to its total occurrences (573). Another distinct misclassification was between Owl (Pecha) and Asian-Koel (Kokil), with 5 such samples being incorrectly predicted.The classification for easily separable birds like Heron (Bok), Coppersmith-Barbet (Basanta-Bouri), Red-Vented-Bulbul (Bulbuli) and White-Rumped-Shama Shama is very high, almost accurate or optimum with least misfiring .Leveraging CBAM (Channel and Spatial Attention Mechanism) also made the model able to concentrate on salient features, especially when distinguishing between difficult bird species with obscure visual resemblance. Performance of the model is confirmed by comparing with conventional CNN architectures in Table 2, which have been proven to provide good accuracies and F1-score, showing superior results in comparison, allowing us to conclude that the approach presented here contributes positively to species classification of birds.

4.1 Explainable AI(XAI) Results and Model Interpretability

The Grad-CAM visualizations in Figure 3 explain the decision of the model, showing which parts of the image affected its predictions. For the Proper class prediction, as in Asian-Koel (Kokil) heatmaps, try to peak at significant versatile characteristics found on birds, which are predicted with high probability (up to 0.99). For misclassifications, such as Asian-Koel (Kokil) predicted as Oriental-Magpie-Robin(Doel), the heatmaps show that the model was focused on a less relevant area,which is demonstrated by a lower confidence score (0.89). So also for the case of misclassifying Asian-Koel (Kokil) as Black Drongo (Kalo-Finge), due to attention on non-salient features, confidence was low (0.81).These results indicate that the model can attend to relevant features in making correct predictions and may provide an increased understanding of sources of confusion in misclassifications,which is helpful for future model refinement.

4.2 Comparison With Existing Works

The comparative in Table 2 highlights the superiority of the proposed **EfficientNetB0+CBAM** model, which achieved an accuracy of **99%**. This outperforms existing approaches, including MobileNet+TL (91%), ViT+TL (92.03%), DarkNet-53+ECOC-SVM (94%), and Dynamic Ensemble (90.5%), thereby demonstrating its effectiveness for Bangladeshi bird classification.

5 Discussion

In this study, we presented BanglaBirds-AttnNet,which is a deep-segmented model based on Convolutional Block Attention Module (CBAM) with EfficientNetB0 architecture, for the classification of endangered Bangladeshi Bird species. Our method successfully overcomes the drawbacks of existing approaches,which

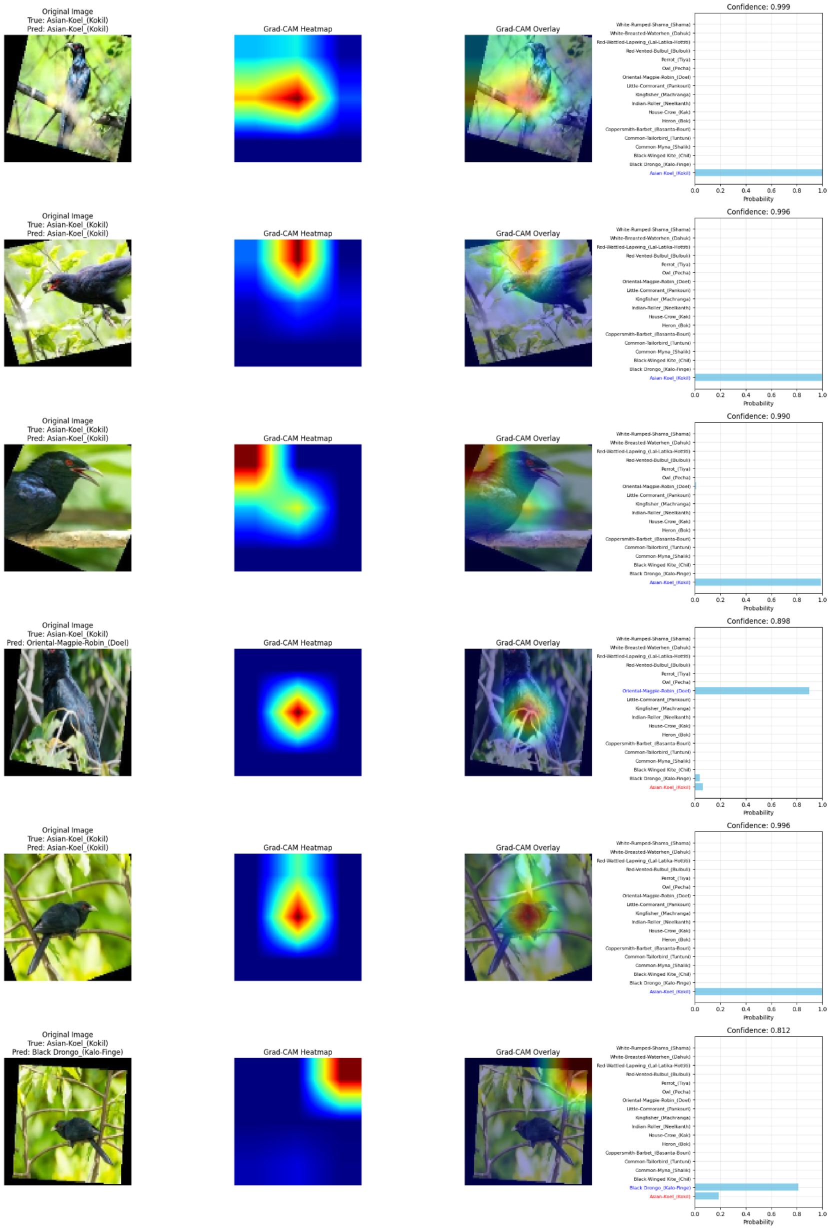

Fig. 3. Grad-CAM visualizations for **Asian-Koel (Kokil)**, showing heatmaps and prediction confidence for correct and misclassified instances.The heatmaps highlight key features used by the model for classification.

Table 2. Performance Comparison of Selected Birds Classification Methods

Author	Method	Year	Accuracy
M. M. Rahman et al. [9]	MobileNet + Transfer Learning	2020	91%
M. H. Rabby et al. [8]	ViT + Transfer Learning	2023	92.03%
Noumida et al. [6]	Res2net + CBAM	2023	93.32%
Qiu et al. [7]	DarkNet-53 + ECOC-SVM	2024	94%
Soumith Gundala et al. [3]	Dynamic Ensemble	2025	90.5%
Proposed Work	**EfficientNetB0 + CBAM**	**2025**	**99%**

are based on generic datasets and fail to capture morphological differences at the fine-grained lexical level. The combination of CBAM allows the model to emphasize important spatial and channel-wise cues, thus improving classification accuracy. A comparison of the experimental results shows that BanglaBirds-AttnNet achieves 99% accuracy and outperforms state-of-the-art models while being computationally efficient. This makes it suitable for accurate classification, which is crucial for the continued monitoring of lead exposure and conservation efforts in Bangladesh. Because we concentrate on vulnerable and endemic birds, our model is of direct relevance to the conservation of avian in Bangladesh. Nevertheless, the current model is constrained by a small dataset (18 species), and in the future, we will consider expanding the dataset's diversity and conducting robustness testing under real-world conditions, such as illumination variance and background noise. The real-time deployment issues, such as edge-device support and process speed, still warrant further investigation. In conclusion, BanglaBirds-AttnNet has enormous potential to aid in Bangladeshi wildlife conservation in addition to advancing bird classification technology. Our study will serve as a starting point for further research in automated biodiversity monitoring and fine-grained recognition.

6 Conclusion

This paper introduces BanglaBirds-AttnNet, a new deep learning architecture that combines the power of EfficientNetB0 with Convolutional Block Attention Module (CBAM) to improve the classification performance of endangered birds in Bangladesh. Our model achieves 99% classification accuracy, demonstrating better performance and efficiency compared to previous methods (based on global data with no attention). We present a framework specifically oriented for real-time, in-the-wild deployment, which can be utilized as an essential tool for biodiversity conservation initiatives in Bangladesh. Future efforts will focus on expanding the dataset size,optimizing the model for improved performance,and enhancing its applicability in larger ecological monitoring scenarios. The performance of BanglaBirds-AttnNet enables more efficient and scalable solutions for wildlife surveillance,which will aid in the conservation of avian biodiversity in Bangladesh.

Acknowledgment

The authors thank the contributors of the BanglaBirds dataset [11] for publishing their data, which greatly helped this work. The authors would like to thank the Department of Computer Science and Engineering,Khwaja Yunus Ali University, for computational resources and academic advice to complete this study.

References

1. Cao, Z., Wang, K., Wen, J., Li, C., Wu, Y., Wang, X., Yu, W.: Fine-grained image classification on bats using vgg16-cbam: a practical example with 7 horseshoe bats taxa (chiroptera: Rhinolophidae: Rhinolophus) from southern china. Frontiers in Zoology **21**(1), 10 (2024). https://doi.org/10.1186/s12983-024-00531-5, https://doi.org/10.1186/s12983-024-00531-5
2. Duc, H., Minh, T., Hong, K., Hoang Luong, H.: 84 Birds Classification Using Transfer Learning and EfficientNetB2, pp. 698–705 (11 2022). https://doi.org/10.1007/978-981-19-8069-5_50
3. Gundala, S., Kanthala, R., Enduri, T.R., Bhaskar, C.V., Kakulapati, V.: Efficient fine-grained bird classification via dynamic ensembles and active learning. Asian Journal of Research in Computer Science **18**(6), 161–168 (May 2025). https://doi.org/10.9734/ajrcos/2025/v18i6688, https://journalajrcos.com/index.php/AJRCOS/article/view/688
4. Lin, T.Y., RoyChowdhury, A., Maji, S.: Bilinear cnn models for fine-grained visual recognition. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 1449–1457 (2015). https://doi.org/10.1109/ICCV.2015.170
5. Lv, D., Zhang, Y., Lv, D., Lu, J., Fu, Y., Li, Z.: Combining cbam and iterative shrinkage-thresholding algorithm for compressive sensing of bird images. Applied Sciences **14**(19) (2024). https://doi.org/10.3390/app14198680, https://www.mdpi.com/2076-3417/14/19/8680
6. Noumida, A., Mukund, R., Nair, N.M., Rajan, R.: Stacked res2net-cbam with grouped channel attention for multi-label bird species classification. In: 2023 31st European Signal Processing Conference (EUSIPCO). pp. 446–450. IEEE (2023)
7. Qiu, Z., Zhou, Z., Wan, Z.: Automatic classification of bird species related to power line faults using deep convolution features and ecoc-svm model. IET Generation, Transmission & Distribution **18**(19), 3138–3149 (2024). https://doi.org/https://doi.org/10.1049/gtd2.13265, https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/gtd2.13265
8. Rabby, M.H., Ahmed, K., Rady, F.R., Akash, A.K.: Birds-bd: A comprehensive dataset for bangladeshi local birds and classification using vision transformer with transfer learning. In: 2023 26th International Conference on Computer and Information Technology (ICCIT). pp. 1–4 (2023). https://doi.org/10.1109/ICCIT60459.2023.10441022
9. Rahman, M.M., Biswas, A.A., Rajbongshi, A., Majumder, A.: Recognition of local birds of bangladesh using mobilenet and inception-v3. International Journal of Advanced Computer Science and Applications **11**(8) (2020). https://doi.org/10.14569/IJACSA.2020.0110840, http://dx.doi.org/10.14569/IJACSA.2020.0110840
10. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV). pp. 618–626 (2017). https://doi.org/10.1109/ICCV.2017.74

11. Shihab, T., Rabeya, T.: Native birds of bangladesh (2024). https://doi.org/10.34740/KAGGLE/DSV/7451671, https://www.kaggle.com/dsv/7451671
12. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. ArXiv **abs/1905.11946** (2019), https://api.semanticscholar.org/CorpusID:167217261
13. Teterja, D., GARCIA-RODRIGUEZ, J., Azorin-Lopez, J., Sebastián-González, E.: A performance evaluation of lightweight deep learning approaches for bird recognition (Sep 2023). https://doi.org/10.1007/978-3-031-43085-5_26, https://doi.org/10.1007/978-3-031-43085-5_26
14. Wang, K., Yang, F., Chen, Z., Chen, Y., Zhang, Y.: A fine-grained bird classification method based on attention and decoupled knowledge distillation. Animals **13**(2) (2023). https://doi.org/10.3390/ani13020264, https://www.mdpi.com/2076-2615/13/2/264
15. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European Conference on Computer Vision (ECCV) (September 2018)
16. Yao, F., Deng, N., Wang, X.a.: Bird recognition based on mixed convolutional neural network. In: Barolli, L. (ed.) Innovative Mobile and Internet Services in Ubiquitous Computing. pp. 235–246. Springer Nature Switzerland, Cham (2024)
17. Yi, X., Qian, C., Wu, P., Maponde, B.T., Jiang, T., Ge, W.: Research on fine-grained image recognition of birds based on improved yolov5. Sensors **23**(19) (2023). https://doi.org/10.3390/s23198204, https://www.mdpi.com/1424-8220/23/19/8204