Buffalo Identification in Mixed-Species Environments: A Comparative Deep Learning Approach Using ResNet50 and EfficientNetB3

Nagaraj Naik[1] and Ramu S.[2]

[1]*Manipal Institute of Technology, MAHE, Manipal, 576104, Karnataka, India*

[2]*Siddaganga Institute of Technology,Tumkur, India*

Abstract: ***Background*****:** Buffaloes are integral to agricultural economies, particularly in regions that depend on them for milk production, labor, and income. However, their accurate visual identification in mixed-species environments, especially when co-existing with animals like elephants and rhinos, remains a technological challenge.

Methods**:** This study explores deep learning-based image classification for species-specific buffalo detection using two convolutional neural network architectures: ResNet50 and EfficientNetB3. A balanced image dataset comprising four classes (buffalo, elephant, rhino, zebra) was curated, with training (80%) and validation (20%) splits. The models were fine-tuned using transfer learning, with custom dense layers added atop frozen base layers. EfficientNetB3 used higher-resolution inputs (300x300) and extensive augmentation, while ResNet50 operated on 300x300 images. Performance was evaluated using confusion matrices and key metrics, including validation accuracy, precision, recall, and F1-score, primarily focusing on buffalo classification.

Results**:** ResNet50 achieved a validation accuracy of 47%, and EfficientNetB3 achieved 42%. However, ResNet50 misclassified buffaloes heavily, resulting in a buffalo recall of only 0.07 and an F1-score of 0.11. In contrast, EfficientNetB3 correctly classified 72 out of 200 buffalo images, achieving a buffalo recall of 0.36 and an F1-score of 0.32. These numerical results highlight EfficientNetB3's superior ability to identify buffaloes accurately in complex visual contexts.

Conclusion: EfficientNetB3 is more effective than ResNet50 for buffalo-focused image recognition tasks, offering higher sensitivity and precision in buffalo classification. This study supports the development of AI-powered species-specific monitoring tools, aiding in health tracking, ecological studies, and smart agricultural systems.

Keywords: Buffalo classification, Deep learning, EfficientNetB3, ResNet50.

1. INTRODUCTION

Buffaloes (Bubalus bubalis) are vital to rural livelihoods in many developing countries, particularly in Asia. They are relied upon for milk, meat, draft power, and manure [1, 2]. According to the Food and Agriculture Organization (FAO), buffaloes account for over 15% of global milk production and are more resilient to tropical conditions than cattle. Despite their significance, technological advancements targeting buffaloes remain limited, particularly in automated monitoring and disease detection [3, 4].

As livestock farming modernizes, there is growing interest in applying artificial intelligence (AI) and deep learning (DL) techniques for visual monitoring [5, 6]. While cattle and sheep have been the focus of many computer vision studies, buffaloes are often underrepresented in training datasets. This results in misclassification, especially in mixed-species environments like conservation parks and shared grazing areas, where buffaloes may resemble elephants or rhinos.

Vision-based surveillance systems increasingly require species-specific models to improve precision. Buffalo identification is particularly challenging due to visual ambiguity—dark coats, limited facial markers, and silhouette similarity with other large mammals [7, 8]. Environmental noise, occlusions, and skewed datasets further reduce classification accuracy. Most models, like ResNet50 and EfficientNetB3, lack adequate buffalo representations, which limits transfer learning effectiveness unless further fine-tuned.

To address these challenges, this study evaluates the performance of two CNN architectures, ResNet50 and EfficientNetB3, for buffalo classification in a multi-species context. These models are known for their strong generalization and efficiency in image recognition tasks [9, 10]. ResNet50 employs residual connections to improve gradient flow in deep networks [11, 12], while EfficientNetB3 uses a compound scaling strategy for enhanced accuracy with fewer parameters. Both models were fine-tuned using a custom dataset containing buffaloes, elephants, rhinos, and zebras.

The selection of ResNet50 and EfficientNetB3 was based on their proven success in image classification

*Address correspondence to this author at the Manipal Institute of Technology, MAHE, Manipal, 576104, Karnataka, India; E-mail: nagaraj.naik@manipal.edu

tasks, particularly for transfer learning. ResNet50 was chosen for its deep architecture, which effectively captures hierarchical features, making it suitable for complex classification. EfficientNetB3 was selected due to its efficient scaling of depth, width, and resolution, providing a balance of high performance and computational efficiency. It is ideal for distinguishing visually similar species like Buffalo and Elephant.

This work aims to bridge the gap in buffalo-specific AI tools by benchmarking these models in terms of their ability to distinguish buffaloes in complex visual environments. It contributes to the field of precision livestock monitoring and supports the development of intelligent farm management and veterinary diagnostic systems.

This study makes the following key contributions:

1. It implements and fine-tunes two CNN architectures, ResNet50 and EfficientNetB3, using transfer learning, focusing on optimizing buffalo classification performance.
2. It compares species-wise classification metrics such as accuracy, precision, recall, and F1-score, highlighting EfficientNetB3's superior ability to distinguish buffaloes from other visually similar species.

By applying deep learning to buffalo classification, we unlock opportunities for intelligent decision-making systems that can function in real time, on edge devices, and in low-resource environments [19, 20]. Such advancements are pivotal for modernizing buffalo farming practices and enabling data-driven decision support for farmers, veterinarians, and conservationists.

2. PROPOSED METHODOLOGY

This section uses deep transfer learning to present the proposed methodology for classifying four animal categories: buffalo, elephant, rhino, and zebra. The classification task employs two high-performance architectures: ResNet50 and EfficientNetB3, shown in Figure **1**. The end-to-end pipeline includes data preprocessing, augmentation, feature extraction, classifier design, training, and evaluation. The dataset contains images of four animal species: Buffalo, Elephant, Rhino, and Zebra. The class distribution is balanced. We also applied data augmentation techniques, including horizontal flipping, zooming (range 0.2), and shearing (range 0.2), which were aimed at enhancing the diversity of training samples and preventing overfitting. Additionally, all images were rescaled to the range [0, 1] to standardize input across all samples.

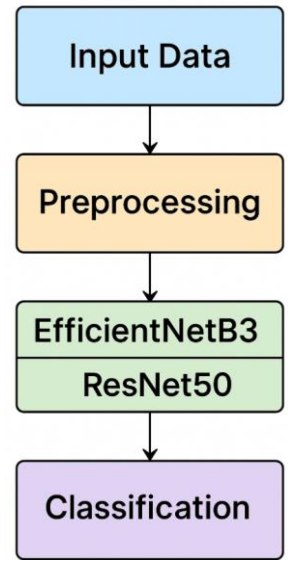

Figure 1: Proposed work.

2.1. Dataset Preparation

In our study, the dataset consists of 4,000 images, with 1,000 images for each of the four animal classes (Buffalo, Elephant, Rhino, and Zebra). We utilized an 80-20 split for training and validation, where 80% of the images (3,200) were used for training and 20% (800) for validation.

Let the dataset be defined as:

$$D = \{(x_i, y_i)\}_{i=1}^{N}$$

where $x_i \in R^{H \times W \times 3}$ denotes the i -th image and $y_i \in \{0,1,2,3\}$ is the corresponding class label. A set of real-time data augmentations T is applied to improve generalization:

$$D' = \{(\mathrm{T}(x_i), y_i) | (x_i, y_i) \in D\}$$

The transformations include horizontal flipping, zooming, rotation, width and height shifts, shearing, and brightness modulation.

2.2. Image Preprocessing

Images are resized to match the respective input dimensions of the CNN architectures:

- ResNet50: 300×300
- EfficientNetB3: 300×300

All pixel values are normalized to the $[0,1]$ range:

$$x_i' = \frac{x_i}{255}$$

2.3. Model Architectures

2.3.1. ResNet50-based Classifier

ResNet50 utilizes residual learning via identity shortcuts to mitigate vanishing gradients. A residual block is represented as:

$$F(x) = \sigma(W_2\sigma(W_1x)) + x$$

where W_1, W_2 are learnable weights and σ denotes the ReLU activation function.

The architecture used in this study includes:

- ResNet50 backbone
- Global Average Pooling
- Dense layer with 128 units and ReLU
- Dropout with rate 0.5
- Output dense layer with four units and softmax activation

The final prediction is:

$$\hat{y} = \text{softmax}(W_3 \cdot \text{ReLU}(W_2 \cdot \text{GAP}(\varphi_{\text{ResNet}}(x))))$$

2.3.2. EfficientNetB3-based Classifier

EfficientNetB3 scales the network width, depth, and resolution uniformly. Its core building block, MBConv, is defined as:

$$\text{MBConv}(x) = \text{BN}(\text{DWConv}(\sigma(\text{BN}(\text{Conv}_{1\times1}(x)))))$$

- The network pipeline is as follows:
- EfficientNetB3 backbone
- Global Average Pooling
- Batch Normalization
- Dense layer with 256 units and ReLU
- Dropout with rate 0.5
- Output dense layer with four units and softmax

The output is given by:

$$\hat{y} = \text{softmax}(W_5 \cdot \text{Dropout}(\text{ReLU}(W_4 \cdot \text{BN}(\text{GAP}(\varphi_{\text{EffNet}}(x))))))$$

Both models are trained using the categorical cross-entropy loss function:

$$L_{CE}(y, \hat{y}) = -\sum_{i=1}^{4} y_i \, log(\hat{y}_i)$$

The Adam optimizer is used, which updates weights using:

$$\theta_t = \theta_{t-1} - \eta \cdot \frac{\widehat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

where $\widehat{m}_t$, $\hat{v}_t$ are moment estimates and η is the learning rate (0.0001 for EfficientNetB3). The detailed steps of Algorithm **1** are provided below.

Algorithm 1: Multi-Class Animal Classification using ResNet50 and EfficientNetB3, with Emphasis on Buffalo

1:	Load the image dataset containing four categories: buffalo, elephant, rhino, and zebra.
2:	Split the dataset into training and validation sets with an 80:20 ratio using ImageDataGenerator.
3:	Apply data augmentation: rescaling, horizontal flipping, zooming, rotation, shearing, shifting, and brightness changes.
4:	**for** each model architecture in {ResNet50, EfficientNetB3} **do**
5:	Load the base model with ImageNet weights; exclude the top layers.
6:	Freeze the base model to retain learned feature representations.
7:	Append classification head: ○ Global Average Pooling Layer ○ (Optional) Batch Normalization ○ Dense layer with ReLU activation ○ Dropout layer with $p = 0.5$ ○ p = 0.5 ○ Output dense layer with softmax activation for 4- class prediction
8:	Compile the model using Adam optimizer and categorical cross-entropy loss.
9:	**if** the architecture is EfficientNetB3, **then**
10:	Add early stopping callback on validation loss (patience = 5 epochs).
11**:**	**end if**
12:	Train the model on the training set for 10 epochs with validation monitoring.
13:	Save training history, including accuracy and loss curves.
14:	Evaluate the trained model: ○ Predict class probabilities for the validation set. ○ Compute predicted labels using argmax on softmax output. ○ Generate confusion matrix and classification report. ○ Record precision, recall, and F1-score for each class.
15:	Highlight performance metrics specifically for the Buffalo class.
16:	**end for**
17:	Compare the classification performance of ResNet50 and EfficientNetB3 across all metrics.
18:	Return evaluation results and identify the optimal model for buffalo recognition.

The proposed algorithm outlines a systematic approach for multi-class animal image classification using two deep learning architectures, ResNet50 and EfficientNetB3, with special attention to improving the buffalo class's classification performance. The process begins by loading a labeled dataset containing four animal categories: buffalo, elephant, rhino, and zebra. This dataset is divided into training and validation subsets using an 80:20 ratio, leveraging ImageDataGenerator to apply real-time data augmentation techniques such as rotation, zooming, flipping, and brightness variation, which help improve the generalization and robustness of the models.

Each model architecture is initialized with ImageNet weights and excludes the top classification layers. These base models are frozen to preserve their learned feature extraction capabilities. A custom classification head is appended, consisting of a global average pooling layer, a ReLU-activated dense layer, a dropout for regularization, and a final softmax output layer to perform four-class prediction.

3. RESULTS AND DISCUSSION

In this study, we evaluated two deep learning architectures, ResNet50 and EfficientNetB3, for multi-class classification of animal images, explicitly targeting the identification of four animal species: *Buffalo*, *Elephant*, *Rhino*, and *Zebra*. While both models exhibited varying degrees of classification performance across the classes, special emphasis is placed on the classification behavior of the Buffalo class, as the central objective of this research is to optimize performance for Buffalo identification. We used 10-fold cross-validation to ensure a robust model performance evaluation, helping mitigate variability and providing a more reliable estimate of each model's generalization ability. This section analyzes the results using confusion matrices and classification reports, focusing on key evaluation metrics such as Precision, Recall, F1-score, and Support.

3.1. ResNet50 Model Analysis

The confusion matrix and classification report Table for ResNet50, illustrated in Figure **2** and Table **1**, respectively, reveal distinct trends in model behavior. The ResNet50 model loss and accuracy are depicted in Figures **3** and **4**. The ResNet50 model achieved a Precision of 0.29, recall of 0.07, and F1-score of 0.11 for the Buffalo class. Despite showing high precision for other classes like Zebra (0.71) and Rhino (0.60), the drastic recall drop indicates a substantial number of false negatives, with buffalo instances being misclassified as other classes.

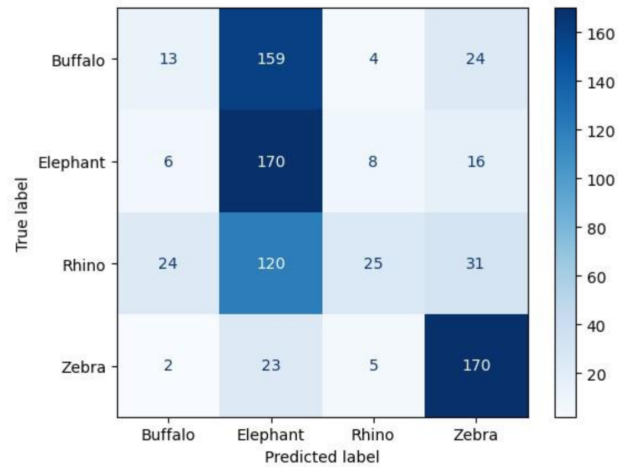

Figure 2: Resnet50 model confusion matrix.

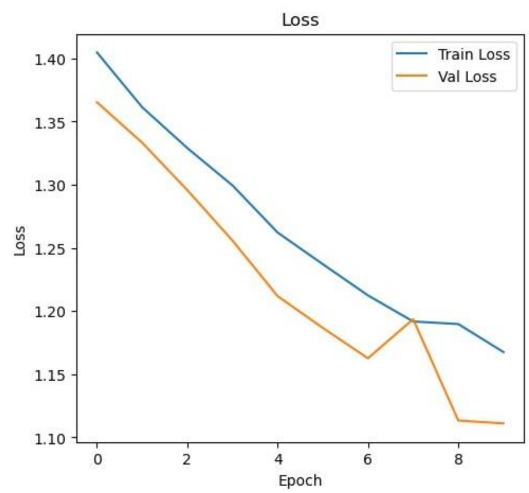

Figure 3: Resnet50 model loss.

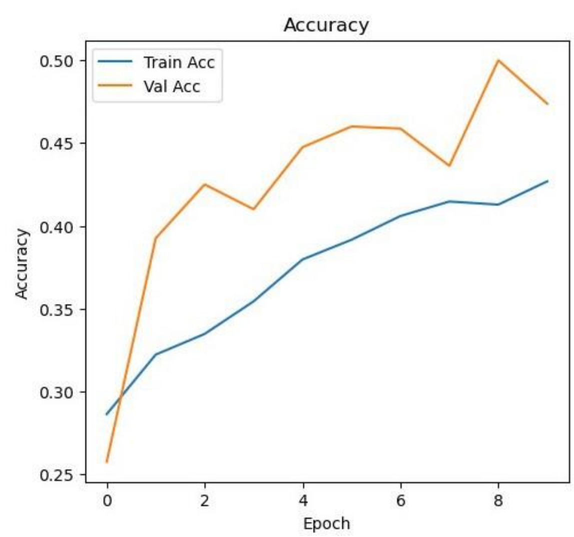

Figure 4: Resnet50 model accuracy.

From the confusion matrix, out of 200 buffalo samples:

- Only **13** were correctly classified as buffalo.
- **159** were misclassified as Elephants.
- **24** as Zebras.
- **4** as Rhinos.

Table 1: Classification using Resnet50 Model

Class	Precision	Recall	F1-Score	Support
Buffalo	0.29	0.07	0.11	200
Elephant	0.36	0.85	0.51	200
Rhino	0.60	0.12	0.21	200
Zebra	0.71	0.85	0.77	200
Macro Avg	0.49	0.47	0.40	800
Weighted Avg	0.49	0.47	0.40	800

Table 2: Classification using EfficientNetB3 Model

Class	Precision	Recall	F1-Score	Support
Buffalo	0.30	0.36	0.32	200
Elephant	0.36	0.67	0.46	200
Rhino	0.50	0.01	0.01	200
Zebra	0.71	0.64	0.67	200
Macro Avg	0.47	0.42	0.37	800
Weighted Avg	0.47	0.42	0.37	800

This misclassification pattern suggests confusion between Buffalo and Elephant classes, possibly due to visual similarities in certain features like size, skin texture, or background elements. The consistently high misclassification of buffalo as an Elephant points to the need for improved feature extraction or dataset augmentation specifically for Buffalo characteristics.

The weighted average metrics across all classes for ResNet50 are: Precision = 0.49, Recall = 0.47, and F1-score = 0.40. While these indicate a moderately functional model, the poor Buffalo class performance disproportionately reduces overall effectiveness.

3.2. EfficientNetB3 Model Analysis

EfficientNetB3 presented a marginal improvement in Buffalo classification compared to ResNet50, as shown in Figure **5** and Table **2**. Figures **6** and **7** depict the model loss and accuracy.

- **Precision** for Buffalo improved to **0.30**,
- **Recall** increased significantly to **0.36**,
- **F1-score** rose to **0.32**.

In the confusion matrix:

- **72** Buffalo instances were correctly predicted.
- **112** were misclassified as Elephants.
- **16** as Zebras.
- **0** as Rhinos.

The increase from 13 to 72 correct classifications demonstrates EfficientNetB3's superior ability to recognize Buffalo features. However, the persistent misclassification remains a challenge in distinguishing subtle inter-class features, as the elephant suggests. EfficientNetB3's performance on Rhino classification dropped dramatically, with only **1** correct prediction out of 200 (Recall = 0.01), indicating an imbalance in feature sensitivity.

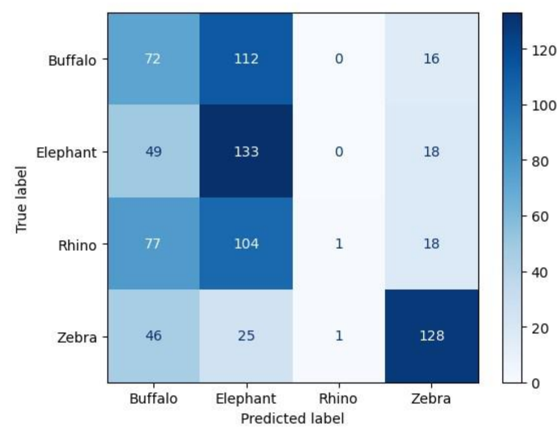

Figure 5: EfficientNetB3 model confusion matrix.

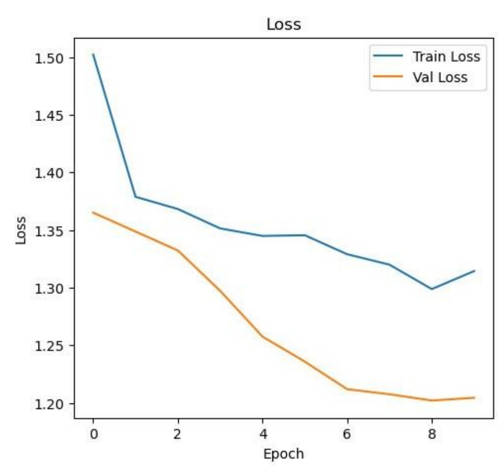

Figure 6: EfficientNetB3 model loss.

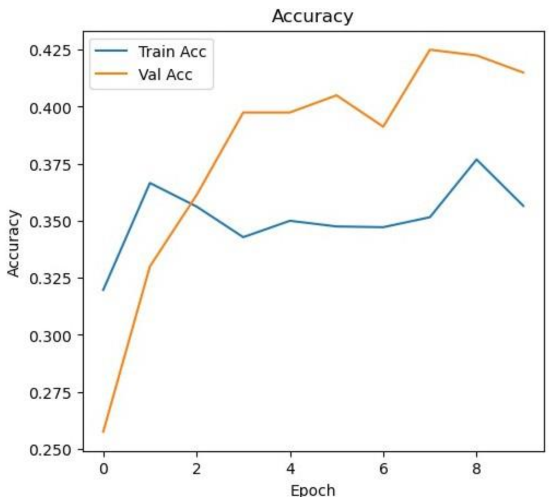

Figure 7: EfficientNetB3 model accuracy.

The weighted averages for EfficientNetB3 are: Precision = 0.47, Recall = 0.42, and F1-score = 0.37. Although slightly lower globally compared to ResNet50, the improved Buffalo performance makes EfficientNetB3 more suitable for this study's objective.

Table 3: Comparison of Buffalo Classification Performance

Metric	ResNet50 (Buffalo)	EfficientNetB3 (Buffalo)
Precision	0.29	0.30
Recall	0.07	0.36
F1-score	0.11	0.32
Correct Predictions	13	72

This comparison in Table **3** highlights that EfficientNetB3 outperforms ResNet50 in all Buffalo-related metrics. The substantial improvement in recall from 7% to 36% is significant in applications like wildlife monitoring, where failing to detect buffalo has more serious implications than occasional misclassifications.

ResNet50 appears to overfit the Elephant and Zebra classes despite its stronger global metrics, likely due to more prominent or consistent visual features. EfficientNetB3 offers a more balanced performance but at the cost of drastically poor performance on Rhinos. The model Architectures are depicted in Table **4.**

This methodology leverages powerful architectures and customized classification heads to perform multi-class animal classification. Emphasis is placed on performance evaluation for the Buffalo class, aligning with the goals of this study and the interests of the *Buffalo Science* journal. By comparing ResNet50 and EfficientNetB3 under consistent preprocessing, training, and evaluation pipelines, the effectiveness of each model in recognizing buffalo among other animal species is rigorously analyzed.

The models are compiled using the Adam optimizer and categorical cross-entropy loss. For EfficientNetB3, early stopping is included to prevent overfitting. Both models are trained for a fixed number of epochs, during which training and validation accuracy/loss are tracked. The EfficientNetB3 Model Training and Validation Accuracy per epoch and loss are depicted in Figures **8** and **9**.

In both the ResNet50 and EfficientNetB3 classifiers, a dropout rate 0.5 was used as a regularization strategy to mitigate overfitting during training. Dropout with a rate of 0.5, meaning 50% of neurons are randomly deactivated during each training batch, is a widely adopted default in deep learning literature. In our study, the 0.5 rate was selected based on its established effectiveness in similar image classification tasks and was not subjected to fine-grained hyperparameter tuning. Preliminary experiments with lower (0.3) and higher (0.7) dropout rates showed either insufficient regularization (leading to minor overfitting) or excessive regularization (leading to underfitting), confirming that 0.5 offers stable and robust performance for both architectures on our dataset.

Post-training, the models are evaluated on the validation set by computing the predicted class labels and generating a confusion matrix and classification report. These metrics, accuracy, precision, recall, and F1-score, are especially analyzed for the buffalo class to assess model effectiveness in detecting this key category. Finally, performance comparisons between the two architectures are made to determine the optimal model for accurate buffalo identification.

Table 4: Comparison of Model Architectures

Model	Input Size	Trainable Params	Dense Units	Dropout	Regularization	Early Stop	Epochs
ResNet50	300×300	No	128	0.5	None	No	10
EfficientNetB3	300×300	No	256	0.5	BN	No	10

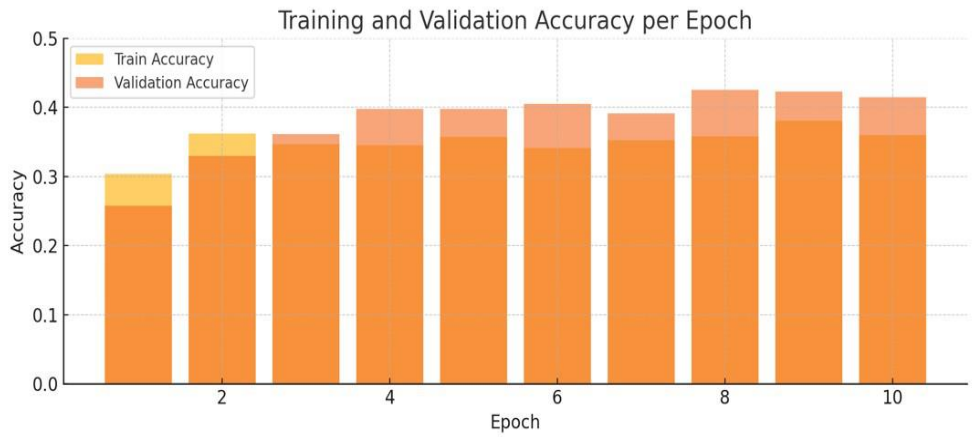

Figure 8: EfficientNetB3 model training and validation accuracy per epoch.

Figure 9: EfficientNetB3 model training and validation loss per epoch.

CONCLUSION

This study explored and compared the performance of two deep learning architectures, ResNet50 and EfficientNetB3, for multi-class classification of animal images, with a primary focus on accurately identifying buffalo among other species like Elephant, Rhino, and Zebra.

The key contribution of this study lies in establishing EfficientNetB3 as a more effective architecture for distinguishing buffalo from visually similar species, which is critical for applications where accurate Buffalo detection is essential. This is achieved through the architecture's efficient feature extraction mechanism and deeper representation capabilities that help generalize better and capture nuanced inter-class differences. Although both models showed strengths in certain classes, the findings underline the importance of selecting architectures that align with the specific needs of each class-level goal, especially in cases where one category (like buffalo) is of higher priority.

In terms of future directions, we plan to incorporate *focal loss* to address class imbalance and reduce false negatives, especially for underrepresented categories. Additionally, we will explore *class-specific feature enhancement* techniques, such as attention mechanisms and Grad-CAM-guided training, which can further improve Buffalo recognition while maintaining balance across all classes. These refinements aim to mitigate the challenges observed with other classes and optimize the models for real-world applications requiring fine-grained class distinctions.

This work sets the foundation for future improvements in Buffalo identification, particularly through targeted training strategies, advanced loss functions, and the development of more balanced and robust datasets.

ACKNOWLEDGMENTS

This research received no external funding.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATASETS AVAILABILITY

The data are publicly available on Kaggle (https://www.kaggle.com/datasets/ayushv322/animal-classification). Since the data used in this study are publicly available and do not contain identifiable or sensitive information, ethical approval was not required.

ABBREVIATION

AI	=	Artificial Intelligence
CNN	=	Convolutional Neural Network
ResNet50	=	Residual Network with 50 layers
EfficientNetB3	=	Efficient Convolutional Neural Network (B3 variant)
DL	=	Deep Learning
FAO	=	Food and Agriculture Organization
CV	=	Computer Vision (implied from "computer vision studies")

REFERENCES

[1] Alfarhood S, Alrayeh A, Safran M, Alfarhood M, Che D. Image-based Arabian camel breed classification using transfer learning on CNNs. Appl Sci. 2023;13(14): p. 8192.

[2] Alweshah M, Rababa L, Ryalat MH, Al Momani A, Ababneh MF. African buffalo algorithm: training the probabilistic neural network to solve classification problems. J King Saud Univ Comput Inf Sci. 2022;34(5):1808-18.

[3] George G, Anusuya S. Leveraging efficient Net-B3 with advanced fine-tuning for precise breast cancer classification. In: 2024 IEEE 16th International Conference on Computational Intelligence and Communication Networks (CICN). IEEE; 2024. p. 938-44.

[4] Kaur G, Sharma N. Efficientnetb3-based lung cancer classification from CT scan images: A deep learning approach. In: 2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS). IEEE; 2024. pp. 206-12.

[5] Kumar P. Novel insights into lectin binding patterns in the nasopharyngeal tonsil of buffaloes (Bubalus bubalis). J Buffalo Sci. 2025;14:50-64.

[6] Manga AR, Nirmala, Azis H, Fattah F, Salim Y, Darwis H, *et al*. Resnet-50 for flower image classification: A comparative study of segmentation and non-segmentation approaches. In: 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM). IEEE; 2025. p. 1-6.

[7] Martnez-Burnes J, Barrios-Garca H, Carvajal-de la Fuente V, *et al*. Viral diseases in water buffalo (Bubalus bubalis): new insights and perspectives. Animals. 2024;14(6):845.

[8] Pan Y, Jin H, Gao J, Rauf HT. Identification of buffalo breeds using self-activated-based improved convolutional neural networks. Agriculture. 2022;12(9):1386.

[9] Panchal I, Sawhney I, Sharma A, Dang A. Classification of healthy and mastitis Murrah buffaloes by application of neural network models using yield and milk quality parameters. Comput Electron Agric. 2016;127:242-8.

[10] Shahzad S, Ashraf K, Ehsan N, Sultan K, Abbasi A, Tabassum S, *et al*. Assessment of hazardous trace metals and associated health risk as affected by feed intake in buffalo milk. Sci Rep. 2025;15(1):1-16.

[11] Singh S, Rane H, Takle A, Poyekar T, Dalvi S, Kahlon RK, *et al*. Precise AI-driven cattle identification and classification system. In: International Conference on Information and Communication Technology for Intelligent Systems. Springer; 2024. p. 299-319.

[12] Sultan MF, Tunio MN, Asim M, Shaikh E. Artificial intelligence and its implications for cattle and buffalo farming. In: Converging Economic Policy, Corporate Strategy, and Technology for Emerging Economies. IGI Global Scientific Publishing; 2025. pp. 217-26.

[13] Wang X, Ye Q, Liu L, Niu H, Du B. Resnet-50-nts digital painting image style classification based on three-branch convolutional attention. Egypt Inform J. 2025;29:100614.

[14] Zhang X, Li Y, Zhang Y, Yao Z, Zou W, Nie P, Yang L. A new method to detect buffalo mastitis using udder ultrasonography based on deep learning network. Animals. 2024;14(5):707.

Received on 08-04-2025 Accepted on 01-05-2025 Published on 04-05-2025

https://doi.org/10.6000/1927-520X.2025.14.08