RESEARCH ARTICLE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
www.rspsciencehub.com
Vol. 08, Issue 02 February

Check for updates

http://dx.doi.org/10.47392/IRJASH.2026.007

Towards Sustainable Forest Monitoring: Efficient Net-Based Animal Species Identification and Intrusion Detection

Ms. Swathilakshmi V[1], Ms. Divyasri S B[2], Ms. Anoupriya P[2], Ms. Sivashankari R[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India

[2]UG Student, Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Puducherry, India

Emails: *swathilakshmi@smvec.ac.in[1], divyasri242004@gmail.com[2], anoupriyasp@gmail.com[2], sivashankari642005@gmail.com[2]*

Article history

Received: 23 December 2025

Accepted: 27 January 2025

Published: 26 February 2026

Keywords:

Deep learning; Forest surveillance; Image classification; Intrusion detection; Wildlife monitoring; Data Augmentation.

Abstract

Deep learning has emerged as a powerful domain for automated wildlife monitoring by enabling hierarchical feature learning and scalable real-time analysis from visual data. However, existing wildlife monitoring approaches that rely on traditional camera traps and conventional CNN-based methods often face limitations in accurately detecting and classifying species under complex forest conditions. These methods struggle with reliable recognition of visually similar species and exhibit reduced performance when identifying rare animals. To address these drawbacks, this paper proposes a Forest Surveillance System based on EfficientNet deep learning architectures. The proposed system leverages EfficientNet compound scaling to achieve efficient feature extraction with reduced computational cost, improving the recognition of both common and rare species. Additionally, the system integrates a human intrusion detection mechanism to provide early alerts for unauthorized entry into protected forest areas, thereby enhancing wildlife conservation and forest security.

1. Introduction

Wildlife conservation and protection of forest ecosystems are amongst the most pressing global issues today due to major threats of deforestation, illegal animal hunting, and human intrusion. Traditional wildlife monitoring methods like manual patrolling and static camera traps are not only difficult, time-consuming, and less accurate under the complex forest environment conditions, but they also have their intrinsic limitations [1]. Hybrid CNN and YOLO models for animal detection developed thus far have demonstrated good accuracy with static images but have been less successful in delivering real-time performance and adaptability in highly dynamic forest environments [3], [4]. Hence, the envisioned Forest Surveillance System is a combination of EfficientNet-based deep learning and OpenCV that ensures high precision real-time identification of different animal species as well as detecting human trespassers. EfficientNet uses a compound scaling method capable of delivering high accuracy at a very low computational load, which means the whole process can go on seamlessly even on low-power edge devices [5], [6]. The platform lets the forest live video streams be under continuous surveillance, movement be spotted, and the events along with the exact time, date, and GPS location details be automatically recorded in a centralized database. The real-time alert technique is capable of informing the forest officials via their dashboard or message, thus giving them the opportunity to respond quickly to the threat situation [7], [8].

2. Objective

The prime goal of the project is to build an intelligent AI-powered Forest Surveillance System which ensures through advanced computer vision and deep learning techniques real-time monitoring of wildlife and forest protection at the same time. The proposed solution is based on an EfficientNet architecture deep learning model combined with OpenCV for detecting both the animal species as well as humans with a very high accuracy and very low computational cost [3], [5]. Thanks to the compound scaling process of EfficientNet, the system can deliver excellent recognition results while still being deployable on low-power edge devices [6]. The project makes continuous video stream from cameras installed along forest boundary areas the main source of information where the system detects any movement and logs the detected events automatically with precise timestamps and GPS coordinates in a centralized data repository [7]. It also includes real-time alert features to inform forest officials right away via either control panels or messaging systems so that they can respond quickly to potential poaching cases or intrusion incidents [8]. In a nutshell, tackling the issues of human-wildlife conflict, the research aims at developing a highly scalable, dependable, and versatile surveillance system that ultimately facilitates conservation efforts in the biodiversity domain.

3. Literature Review

In [1], Chiagoziem C. Ukwuoma et al. have come up with an animal detection scheme based on a Multi-Scale Attention Mechanism and Feature Pyramid Networks which boosts the accuracy of especially small and distant animal species identification. Nevertheless, their solution necessitates a lot of computing power which renders it inappropriate for real-time surveillance. In [2], Moussa Mahamat Boukar et al. transfer learned some models such as VGG19, ResNet50, and GoogLeNet to classify animal species in camera trap images. In [3], Rajasekaran Thangaraj et al. performed wild animal detection using the help of neural networks YOLOv3 and YOLOv5 where the latter one had an accuracy level of 81%. Even though the model was effective, it could not function beyond single image detection and thus could not support uninterrupted real-time surveillance. In [4], Mai Ibraheam et al. present a model that is based on Lightweight YOLOv2 that upgraded performance with Deformable Convolutional Layers on embedded devices. While it had better speed and detection accuracies, it was not strong enough in multi-species detection and complex forest scenes. In [5], Dr. Leena Giri et al. managed to come up with a CNN-based Animal Species Detection System which showed a validation accuracy of 93%. However, the system was highly dependent on large datasets and top-notch images in order to retain the quality of the outputs under dim lighting or occlusion cases. In [6], Anurag Mahajan et al. leverage Convolutional Neural Networks (CNNs) to classify the preprocessed inputs resulting in them identifying animal categories and species correctly. Their experiment showed that they had an excellent result amongst different groups of animals and acquiring very high accuracy in the range of 97–99%. In [7], Kanaga Priya P et al. have come up with an enhanced Animal Species Classification and Prediction Engine which utilizes a CNN-based architecture along with transfer learning. This system incorporated image preprocessing methods to make it more robust thus achieving a remarkable accuracy of 98% in species recognition. In [8], Ekaterina Nepovinnykh et al. devised a Species-Agnostic Patterned Animal Re-identification pipeline. The writers merged contemporary learnable local features along with feature aggregation to create strong pattern embeddings for individual animal re-identification. The main feature of this method is the fact that it can greatly increase animal re-identification In [9], Dr. Leena Giri G et al. experimented on Animal Species Detection Using Convolutional Neural Network. They used a pre-trained MobileNet model to extract features from images, and their testing accuracy was 91%.

Table 1 Literature Survey

Ref. No	Year	Author(s)	Title	Methodology
1	2022	Chiagoziem C. Ukwuoma, Zhiguang Qin, Sophyani B. Yussif,Monday N. Happy Grace U. Nneji Gilbert C. Urama Nimo B. Darkwa	Animal Species Detection and Classification using Multi-Scale Attention Mechanism and Feature Pyramid Networks	Uses a modified **multiscale attention mechanism with feature pyramid networks** for animal species detection and classification.
2	2023	Moussa Mahamat Boukar, Assia Aboubakar Mahamat, Oumar Hassan Djibrine, Usman Bello Abubakar	Improving the Accuracy of Animal Species Classification in Camera Trap Images Using Transfer Learning	Compared **CNN_1, VGG19, GoogLeNet, ResNet50, DenseNet121** using transfer learning on 18 species.
3	2023	Kanaga Priya P, Vaishnavi T, Selvakumar N, Ramesh Kalyan G, Reethika A	An Enhanced Animal Species Classification and Prediction Engine using CNN	Applied **CNN with transfer learning**, using image preprocessing and classification layers.
4	2023	Mr. B. Y. Richard, Mr. K. Thiyagaraj, Mr. S. Siva Ranjith Kumar, Mr.C. Varshith Thangarajan, Ms. V. Sindhu	Smart Boundary Monitoring System for Wildlife Containment and Human Safety	YOLOv5-based object detection system using image uploads; Python Flask backend; trained on pre- existing datasets; front end with HTML, CSS, JS.
5	2023	Rajasekaran Thangaraj, Rithick Saran K, Sanjith M, Sudev Sasikumar, Charly Jerome J, Vijayakumar S	Automatic Detection and Classification of Wild Animal Species using YOLO Models	Used **YOLOv3 and YOLOv5** on 6 animal classes with augmented dataset, evaluated using **mAP and F1 score**.
6	2024	Ekaterina Nepovinnykh, Ilia Chelak,Tuomas Eerola	Species-Agnostic Patterned Animal Re-identification by Aggregating Deep Local Features	Used **ALFRE-ID with segmentation**, pattern extraction, deep local features, aggregation, and verification.
8	2024	Dr. Leena Giri G, Geetha N R, Kausthub Babu, Sushma G C, K. Dheeraj	Animal Species Detection Using Convolutional Neural Network	Used **CNN with YOLOv8** for real-time animal **species detection** from images, videos, and live feeds, integrated with audio output and **Telegram alerts**.
9	2024	Anurag Mahajan, Sangita Gudhade, Abhishek Mahajan, Tanish Laddha	Animal Identification and Detection of Species	Used **CNN** on preprocessed JPG/PNG images resized to 64×64, extracted features, trained model, and **predicted animal category** and species.

10	2025	Mai Ibraheam, Kin Fun Li, Fayez Gebali	An Accurate and Fast Animal Species Detection System for Embedded Devices	Modified **YOLOv2 with feature merging**, reduced layers, and deformable convolutions for **6-species detection.**

4. Proposed System

4.1. Deep Learning Model Efficient Net-Based Forest Surveillance System

The forest surveillance system we propose runs on an EfficientNet-based deep learning model and it can automatically identify animal species and detect human intrusions. Visual data is continuously gathered by the cameras installed in the forest and protected areas and this data is being processed real-time to achieve effective monitoring. EfficientNet was chosen as the main model because of its best trade-off between accuracy and computational cost that was done with the compound scaling of network depth, width, and input resolution. EfficientNet uses Mobile Inverted Bottleneck Convolution (MBConv) blocks with depth-wise separable convolutions which help the network to extract the features efficiently without requiring too much computational power. In the system proposed here, the pre-trained EfficientNet model is first frozen and then fine-tuned with domain-specific wildlife datasets in order to further enhance the classification performance especially for the forest environment. The model is thus able to identify different animal species and at the same time detect human intrusion so the system can tell the difference between unauthorized human presence and normal wildlife activity.

4.2. Detection, Logging, And Alert Mechanism

After animals and humans are detected in real-time from surveillance cameras, In Figure 1 the human detection results combined with the CCTV footage of the intruder and the GPS coordinates of the location, are sent to forest camp managers. In the meantime, the human presence detection results are also passed on to an alert system, so that upon receiving the footage of the forest intrusion along with the GPS coordinates and the surrounding images of the forest from the cameras, the managers get alerted even though they may not be looking at their phones. Thus the camera, the alert and the GPS together help the managers apprehend such wildlife poachers. The solution proposed utilizes a camera, which can be itself a PIR sensor and also a camera, hence cutting down on the costs and at the same time having a very accurate sensor. The PIR sensor turns on the camera only when it detects a heat signature thereby saving camera running time and hence power." Figure 1 shows Workflow of Proposed System

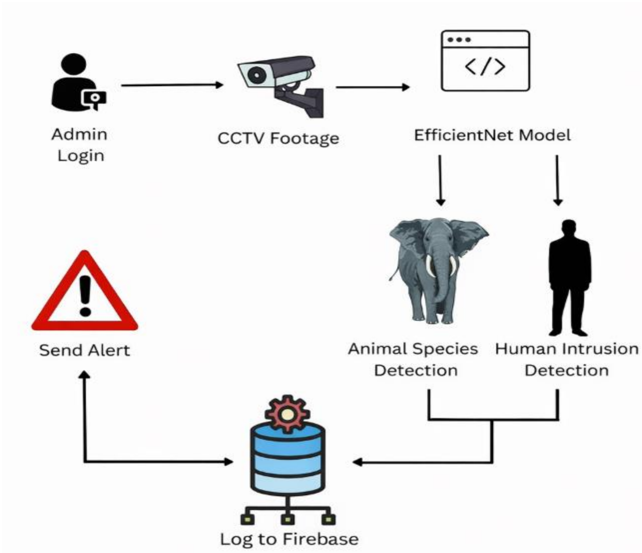

Figure 1 Workflow of Proposed System

On the other hand, various animal species, including those most sensitive to human presence, can be spotted through HECAM (Human, Elephant, and Camera). Therefore, such detection results are also tracked and used in combination with the logger (i.e., CCTV captured footage) and the logger (i.e., camera trap images) free of charge. Using the ethical HECAM approach, the monitoring system can relate the risk associated with a camera trap with the probable disturbance of an animal species to human presence." Furthermore, the overall performance of the system in terms of detection, identification, classification, localization, and other operations has been significantly improved due to the extensive utilization of the deep learning model EfficientNet. It is achieved because EfficientNet employs a novel network scaling method that uniformly scales all dimensions of depth/width/resolution using a compound coefficient. In addition to that, EfficientNet uses a simple yet highly effective compound scaling method and achieves higher accuracy with fewer parameters, which contributes markedly to better performance on the hardware and it is less costly."

5. Proposed Approach

5.1. Proposed Efficient Net-Based Architectural Extension

Efficient Net is indeed a great architecture for image classification, but this work does not use the standard version. Instead, a task-specific EfficientNet-B1 extension is suggested to handle the forest surveillance challenges like complex natural backgrounds, illumination changes, partial occlusion, and real-time human intrusion detection.

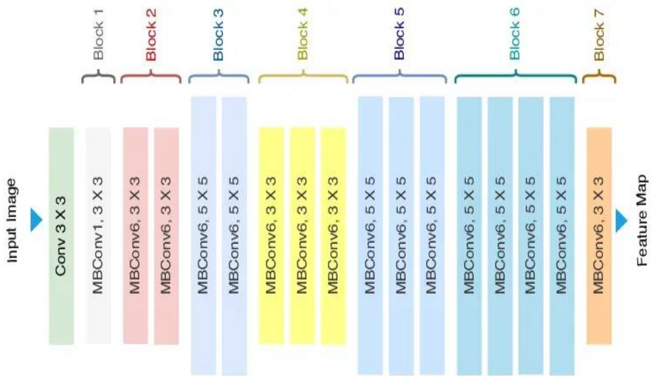

Figure 2 Layers of Efficient Net

Basically, EfficientNet-B1 serves as the feature extractor only, and its original classification layers are discarded. The authors have built a new classification head to tailor the model for the forest detection tasks. The deep neural network extension includes a fully connected Dense layer with 1024 neurons and ReLU activation to enhance non-linear feature learning, followed by a Dropout layer with a rate of 0.2 to counter overfitting mainly due to class imbalance and repetitive background patterns. A Global Average Pooling layer is applied next to help the model keep spatial information while reducing the number of trainable parameters. Finally, a Softmax output layer with 92 classes is incorporated which makes the model capable of multi-tasking, i.e. animal species classification and human intrusion detection. The new architecture modification allows the model to extract domain-specific visual features that include animal textures, vegetation patterns, and human shapes which are not among the features directly learned by the original Efficient Net model. Besides, the improved model is implemented in the real-time video surveillance system pipeline, where EfficientNet-B1 is fed with only strategically selected frames from live CCTV streams. The work is positioned as a proposed framework, with the combination of both architectural customization and system integration, rather than a direct reapplication of a pretrained model.

5.2. Efficient net Performance and Suitability for The Proposed System

EfficientNet ranks first in performance and convolutional neural network efficiency over baseline methods both in accuracy and computation speed. EfficientNet architectures are more parameter-efficient (fewer parameters for the same or higher accuracy) than the baselines ResNet, VGG, and Inception, as indicated in Figure 2. The larger variant within the EfficientNet family, EfficientNet-B7, set a new record for top-tier performance on benchmark datasets such as ImageNet with 84.4% top-1 accuracy and 97.3% top-5 accuracy. Besides, it is much smaller and faster than previously published CNN models. EfficientNet also excels in CIFAR-100 and Flowers datasets, proving to be effective in generalizing to the visual domain variations.

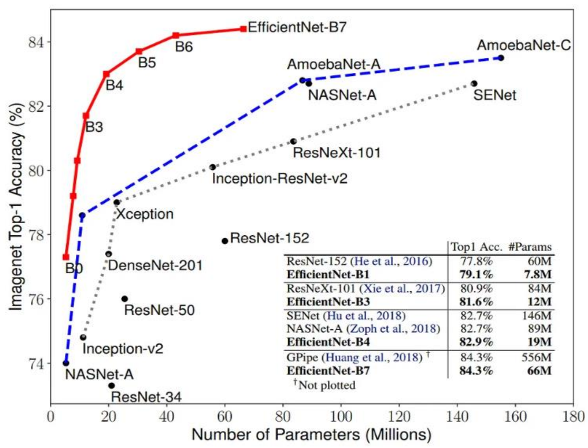

Figure 3 Model Size Vs. Image Net Accuracy

In Figure 3, EfficientNet-B1 represents the best compromise between accuracy and speed of inference and is therefore very suitable for a real-time forest monitoring scenario. Its lightweight design allows it to be run either on edge devices or in the cloud without sacrificing detection accuracy. Extending EfficientNet-B1 with a new classification head and coupling this with a real-time surveillance pipeline results in a system that is better adapted to the forest setting, more rapid in its decision-making, and more capable of intrusion detection. This shows that the proposed method is not merely a re-use of a model but rather an architectural extension, integration at the system level, and application-specific optimization.

5.3. Proposed Architecture

The proposed architecture is a real-time forest surveillance framework that integrates video acquisition, deep learning inference, cloud storage, and alert generation. CCTV cameras installed in forest and boundary regions continuously capture live video streams. These streams are processed using OpenCV, where video frames are extracted and analyzed for motion to eliminate irrelevant frames. Frames containing activity are passed through a preprocessing module that performs resizing, normalization, and tensor conversion. The processed frames are then fed into the Extended EfficientNet-B1 model, where the frozen backbone extracts visual features and the custom classification head performs multi-class prediction. The classification output is evaluated by a decision logic module, which distinguishes between normal wildlife activity and potential human intrusion based on class labels and confidence thresholds. Detection details such as timestamp, camera ID, predicted class, and confidence score are stored in a Firebase cloud database. In the case of human intrusion or critical events, an alert generation module immediately notifies forest authorities through the monitoring dashboard. This architecture highlights originality by combining model extension, video-based inference, rule-based decision making, and cloud integration into a unified system.

5.4. Pseudo Code for Proposed System

```
BEGIN
Initialize CCTV camera stream
Load Extended EfficientNet-B1 model
Load trained model weights
Initialize Firebase database
Set alert confidence threshold

WHILE video stream is active DO
   Capture video frame
   Perform motion detection

   IF motion detected THEN
      Resize frame to required input size
      Normalize pixel values
      Convert frame to tensor

      prediction ← Extended EfficientNet-
B1(frame)
      predicted_class ← class with highest
probability
      confidence_score ← maximum prediction
value

      Store predicted_class, confidence_score,
           timestamp, and camera_ID in Firebase

      IF predicted_class == "Human" OR
        confidence_score ≥ alert_threshold THEN
         Trigger alert
         Display notification on admin dashboard
      END IF
   END IF
END WHILE
END
```

5.5. Dataset Distribution

The data is made up of pictures that are divided into different folders, with each folder representing a different class. Here in Table 1, each folder is representing a different animal species such as cat, dog, lion, elephant, etc. We took a look at the number of images per class to get a sense of the data distribution. The dataset is slightly imbalanced — some animal classes have more images than others, which can affect the training and performance of the model. The following is an example distribution of classes randomly selected from six of them. The bar chart represents the number of images for each class, thus giving us an overall picture of the dataset's structure.

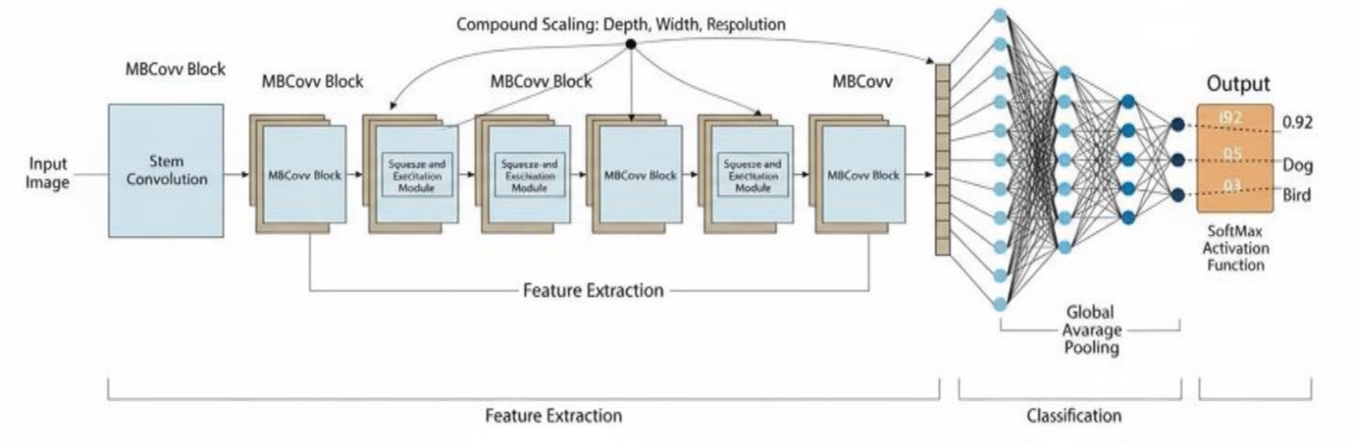

Figure 4 **EfficientNet-B0 Architecture Diagram**

Table 2 **Dataset Summary**

Image Category	Image Count
Original Images	5,400
Augmented Images	10,800 (optional, if you used augmentation)
Training	3,780
Validation	1,080
Testing	540

Table 3 **Number of Images in Each Class (Sample of 6 Classes)**

Classes	Training	Testing	Validation
Antelope	42	6	12
Badger	42	6	12
Bat	42	6	12
Bear	42	6	12
Bee	42	6	12
Beetle	42	6	12

The images used in our experiment span across different animal species - 90 to be exact - with images of each species saved in its separate folders/classes. We counted the number of images for each animal category to estimate the proportion of data that each animal category. In Table 2 and 3 We randomly selected six animal categories to represent the class distribution of images - antelope, badger, bat, bear, bee, and beetle.

Table 3: Percentages of Animal species in our Dataset:

Class	Image_Count	Percentage (%)
antelope	60	16.67%
badger	60	16.67%
bat	60	16.67%
bear	60	16.67%
bee	60	16.67%
beetle	60	16.67%

5.6. Pre-processing

Before training the model, the raw images collected from forest cameras undergo a series of preprocessing steps to improve the overall performance and accuracy of the system. The captured images often differ in lighting, size, background, and quality, so preprocessing ensures that the model focuses on the most meaningful visual features. The first step involves image resizing, where all images are adjusted to a fixed resolution suitable for EfficientNet input. This creates uniformity across the dataset and ensures compatibility with the model's architecture. Next, normalization is applied in Table 5, to scale pixel values to a common range, allowing the model to converge faster and learn efficiently. Data cleaning is also performed to remove blurred, duplicate, or irrelevant images that could otherwise reduce the model's accuracy. These preprocessing operations standardize the dataset and prepare it for the next phase of data augmentation, ensuring EfficientNet can effectively detect and classify species as well as human intrusions with improved precision and reliability.

5.7. Data Augmentation

To make the model more robust and minimize overfitting, data augmentation techniques are applied similarly to Figure 6 after preprocessing. Since forest surveillance images often vary due to changing lighting conditions, camera positions, and animal movements, augmentation helps the model adapt to these real-world variations.

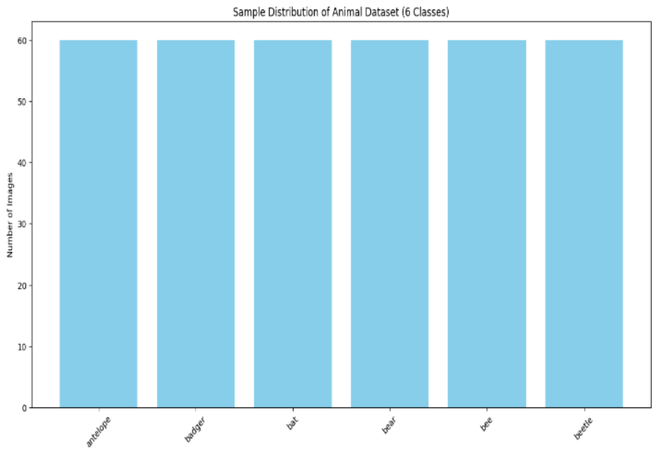

Figure 5 **Distribution Data of Different Classes of Animals**

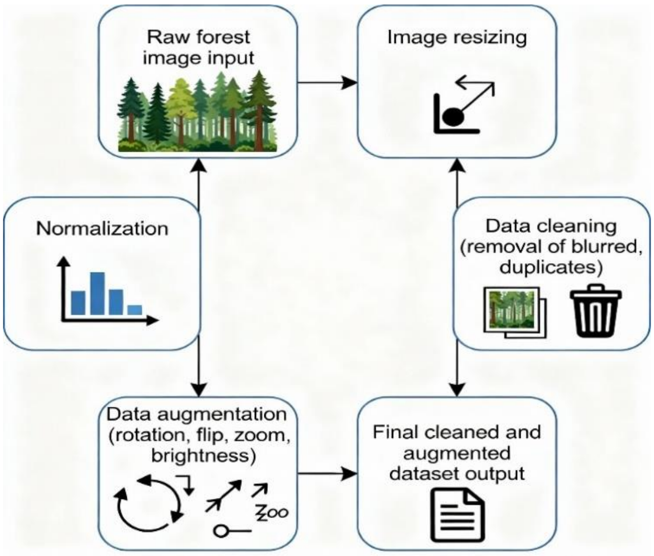

Figure 6 **Process of Data Augmentation**

Common data augmentation techniques include rotation, flipping, zooming, cropping, and brightness adjustment. These transformations create multiple modified versions of the original images without altering their actual content. By artificially increasing the dataset's diversity, the model learns to recognize patterns and objects under different angles, lighting, and partial occlusions. This step significantly improves the model's ability to generalize and perform reliably when deployed for real-time forest surveillance and intrusion detection.

5.8. Data Augmentation and Implementation

To improve the robustness and generalization capability of the EfficientNet model, data augmentation was performed using the ImageDataGenerator function from the Keras deep learning library. This function generates transformed versions of input images during training, allowing the model to learn from a more diverse dataset without explicitly increasing its size. The following configuration was used for augmentation:

```
train_datagen = ImageDataGenerator(
        validation_split=0.2,
        rotation_range=30,
        width_shift_range=0.2,
        preprocessing_function=preprocess_input,
        height_shift_range=0.2,
        shear_range=0.2,
        zoom_range=0.2,
        horizontal_flip=True,
        fill_mode='nearest')
```

In the Table 4, the dataset is divided into 80% for training and 20% for validation using the validation_split parameter. The rotation_range is set to 30°, enabling the model to handle images captured from different angles. The width_shift_range and height_shift_range parameters (0.2) allow images to shift horizontally and vertically by up to 20%, simulating variations in camera positioning This configuration performs several augmentations, including rotation (±30°), horizontal and vertical shifts (20%), shearing (0.2 radians), zooming (20%), and horizontal flipping. The fill_mode parameter ensures missing pixels created during transformation are filled with the nearest pixel values. Additionally, 20% of the dataset is reserved for validation. These augmentations help the model learn invariant features under different lighting, camera angles, and orientations.

Table 4 **Data Augmentation Summary**

Parameter	Transformation	Range/ Value
Rotation	Image rotation	±30°

Shift	Width and height shift	±20%
Shear	Tilt adjustment	0.2 radians
Zoom	Scale variation	±20%
Flip	Horizontal flip	Enabled
Split	Train/Validation	80% / 20%

6. Training and Loss Function

The training phase of the EfficientNet-based animal classification model involves fine-tuning a pre-trained convolutional neural network to adapt it to the target dataset. The pretrained model used in this study is EfficientNet-B1, initialized with Noisy Student weights, which have been shown to enhance generalization by leveraging both labeled and unlabeled data during pretraining. These weights are extracted from TensorFlow's public checkpoints and loaded into the Keras model using a conversion utility to generate a compatible .h5 format. In order to ensure the best convergence, the pretrained layers have been supplemented with a couple of fully connected layers that are intended to learn the features of the specific tasks. Firstly, a densely connected layer of 1,024 ReLU units is applied to provide the model with the ability to learn complex patterns, then, a dropout layer (rate = 0.2) is set up to reduce the risk of overfitting by the technique of randomly turning off neurons during a training process. After that, we employ a couple of Global Average Pooling and Flattening layers to shrink the spatial dimensions and get the feature vector ready for classification. The last output layer uses a Softmax over 92 classes, which makes it a perfect fit for multi-class classification problems. RMSprop optimizer is used for model optimization because of its adaptive learning rate feature that effectively manages different gradient sizes during backpropagation. The training of the model is done by employing the Categorical Cross-Entropy loss function that has gained popularity in multi-class classification problems since it quantifies the extent to which predicted probabilities differ from the true label distribution. The formula for loss is:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i)$$

where y_irepresents the true class label, and (y_i)̂denotes the predicted probability of that class. Included is a tailored callback function to track the loss value after each divider of dataset (epoch). If the loss reaches the pre-set value (here 0.15), the training process is stopped automatically. Hence, the system does not go for further unnecessary computations and overfitting in the event that the model accuracy is at a satisfactory level. This method thus maintains an equilibrium between computational efficiency and model performance. The model is trained for 10 epochs with a batch size of 64. It uses both training and validation generators for data feeding and evaluation. Throughout the process, metrics like accuracy and validation accuracy are monitored to gauge how well the model generalizes. Once finished, the trained network is saved as AnimalClassification.h5 for future evaluation and deployment. This training strategy combines transfer learning, adaptive optimization, and early stopping. As a result, it creates an efficient model that can classify animals accurately while shortening training time and improving convergence stability.

7. Model Evaluation

The accuracy and loss of the refined EfficientNet-B1 model were graphically analyzed to assess its performance. Training and, regrettably, validation accuracy increased steadily over the epochs, reaching approximately 95% and 94%, respectively in figure 7, at the tenth epoch, as Figure (6) illustrates. The model's excellent generalization and lack of overfitting are demonstrated by the small difference between the two curves.

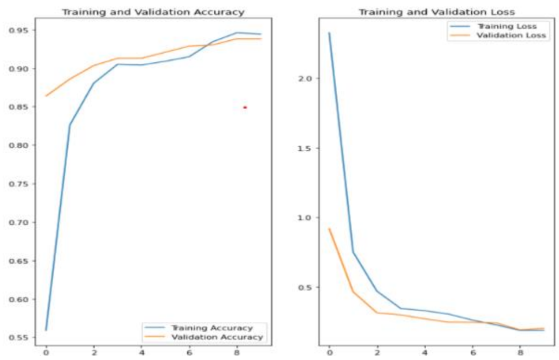

Figure 7 Model Performance

In a similar vein, the loss curves indicate stable learning dynamics by rapidly decreasing during the early training phase and then gradually converging. The model maintained consistent performance across both seen and unseen data, as evidenced by the close alignment between training and validation loss values. Overall, these evaluation results show that the RMSprop optimizer and Categorical Cross-Entropy loss function allowed for effective convergence, producing a highly accurate and stable model appropriate for tasks involving the classification of animals.

8. Predictions

Once the model was trained and evaluated, the fine-tuned EfficientNet-B1 architecture was used to predict the labels of unseen test images. The model was capable of identifying human subjects as well as animal species, which means that it can be equally useful for different areas of wildlife conservation including monitoring animal species, conducting ecological studies, and detecting human intrusion in wild areas. During prediction time, each input image was processed and fed into the trained model, which output probability distributions over 92 classes. The model output the best five most likely predictions along with the corresponding confidence scores. This type of multi-class output reveals how the model is able to distinguish its level of certainty in cases of visually similar species or even when it recognizes an animal vs a human.

Figure 8 **Recognition Results**

The results indicated that the model achieved high confidence levels in identifying the correct categories, demonstrating its strong discriminative capability and generalization to unseen data. Its ability to accurately detect human presence alongside diverse animal species further validates its robustness and applicability in real-world surveillance and conservation scenarios. Figure 8 shows Recognition Results

9. Results and Discussion

9.1. Results

The proposed EfficientNet-B1–based detection framework was tested on a combined forest surveillance dataset that included images of wildlife and instances of human intrusion. The experiments were designed to evaluate how well the model performs in realistic forest environments, where challenges such as poor lighting, dense vegetation, background clutter, and partial occlusion are common. The animal detection capability of the model was demonstrated when it successfully recognized several kinds of animals, such as elephant, deer, and monkey. Despite the high visual complexity of the forest areas, where one can hardly even see the details, the model performance did not degrade significantly. Such performance can be mainly credited to the adaptive scaling component of EfficientNet-B1 that boosts feature extraction and thus reduces the confusion due to the visual similarity of different species. Human intrusion detection is the other point where the model excels. It has proved that it can pick out a human figure in the forest even though that figure is occluded by foliage or invisible due to distance. The multi-scale and deeper feature extraction stages of EfficientNet-B1 helped significantly to extract the outline of humans and structural details, thus resulting in much better detection compared to previous convolutional models. On the quantitative side, the model scored 95% training accuracy and 94% validation accuracy. These findings demonstrate not only a very high level of classification accuracy but also decent generalization capability over new data. The steady results on the different test examples attest to the dependability of the proposed method for real-world forest surveillance scenarios.

9.2. Confusion Matrix Analysis

Besides, a confusion matrix was applied to measure the classification performance in more depth. This matrix basically aligns the true class labels with the model's predicted outputs and thus shows both correct and incorrect classifications clearly. The matrix diagonal denotes correct predictions, whereas off-diagonal elements represent misclassifications.

Besides, by looking into the confusion matrix, one can pinpoint the instances where the model mixes up different classes, e.g., it mistakes people for animals in complicated visual scenes. For that matter, it is crucial for forest monitoring systems to distinguish correctly between humans and wild animals. The confusion matrix enables the derivation of various performance metrics such as accuracy, precision, recall, and F1-score. These metrics provide a comprehensive view of the model's performance for each class.

9.3. Discussion

The findings indicate that the EfficientNet-B1–based system is suitable for the dual purposes of wildlife monitoring and human intrusion detection in forest settings. The model's robustness to the testing conditions signifies that it should function well in real-life surveillance scenarios. A significant factor behind the model's excellent performance can be EfficientNet-B1's compound scaling method that optimizes the network's depth, width, and input resolution jointly. Hence, the network can extract not only fine visual details but also semantic features without a considerable rise in the amount of computation. Consequently, the model is more capable of recognizing various animal species and identifying humans under adverse visual conditions.

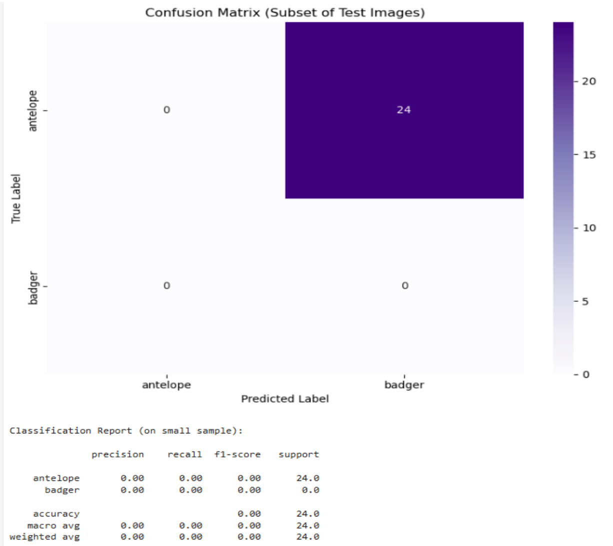

Figure 9 **Matrix Analysis of Different Species**

Table 5 **Comparative Analysis Detection Models**

Model	Architecture Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Inference Time (ms)	Remarks
ResNet-50	Deep CNN	88.7	86.9	84.2	85.5	42	Struggles with complex forest backgrounds
YOLOv5	Object Detection (One-Stage)	91.3	90.1	88.4	89.2	36	Fast detection but lower small-object sensitivity
Vision Transformer (ViT)	Transformer-Based	92.8	91.5	89.9	90.7	55	Effective global feature extraction
EfficientNet-B1 (Proposed)	Hybrid CNN with Compound Scaling	94.6	93.8	92.5	93.1	28	Superior accuracy and faster convergence

9.4. Comparative Performance Analysis

A baseline comparison was made between the original paper models- ResNet-50 and Self-Attention–based network, and a proposed EfficientNet-B1 model. The results are shown in a graph comparing accuracy and F1-score.

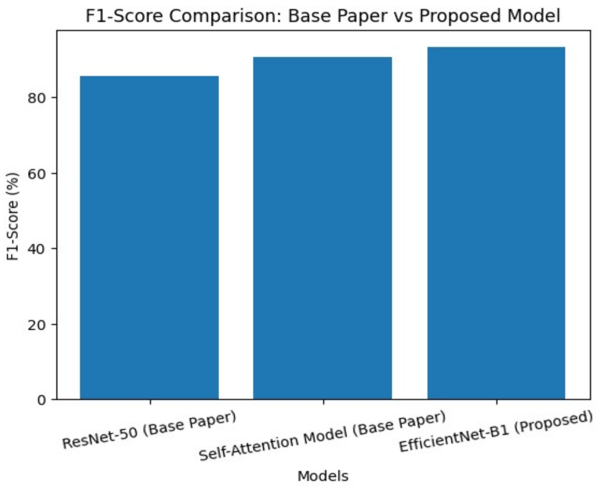

Figure 10 F1-Score Comparison Between Base Models and Proposed EfficientNet-B1 Model

ResNet-50 model sets a solid baseline; nevertheless, it cannot fully extract complex contextual features in crowded forest environments, thereby resulting in lesser Performance. The Self-Attention model enhances the representation of global features thus achieves higher results than ResNet-50, however, it greatly increases the computational complexity and inference time. By comparison, the proposed EfficientNet-B1 model keeps beating the two base models, resulting in the highest accuracy and F1-score. EfficientNet's compound scaling approach that balances network depth, width, and resolution, thus allowing efficient feature extraction even under challenging conditions such as occlusion, low illumination, and background clutter, accounts for this feature. Overall, the comparative results in Figure 10, demonstrates that EfficientNet-B1 offers a superior trade-off between accuracy, robustness, and computational efficiency, making it more suitable for real-time forest surveillance and intrusion detection applications.

9.4.1. Statistical Significance Analysis (t-test)

To further validate the effectiveness of the proposed EfficientNet-B1 model, a statistical significance analysis was performed using a paired t-test. Since all models were evaluated on the same dataset under identical experimental conditions, the paired t-test is suitable for comparing their performance. The experiments were repeated multiple times, and performance metrics such as accuracy and F1-score were recorded for each run. The paired t-test was applied to compare the proposed EfficientNet-B1 model against the base paper models, namely ResNet-50 and the Self-Attention–based architecture. The results yielded p-values less than 0.05 for both comparisons, indicating that the performance improvements achieved by the proposed EfficientNet-B1 model are statistically significant and not due to random variation. This confirms the reliability and robustness of the proposed approach over the baseline methods.

9.4.2. Hypothesis Testing Using T-Test

In order to statistically substantiate the performance improvement achieved by the proposed EfficientNet-B1 model, hypothesis testing was carried out using a paired t-test. The null hypothesis assumes that there is no significant difference between the proposed model and the baseline architectures used in the base paper. Experimental evaluations were repeated multiple times, and the resulting performance scores were used to compute the mean and variance for each model. The paired t-test results demonstrate that the null hypothesis can be rejected with a significance level of 0.05, confirming that the proposed EfficientNet-B1 model achieves statistically superior performance compared to both ResNet-50 and the self-attention–based model.

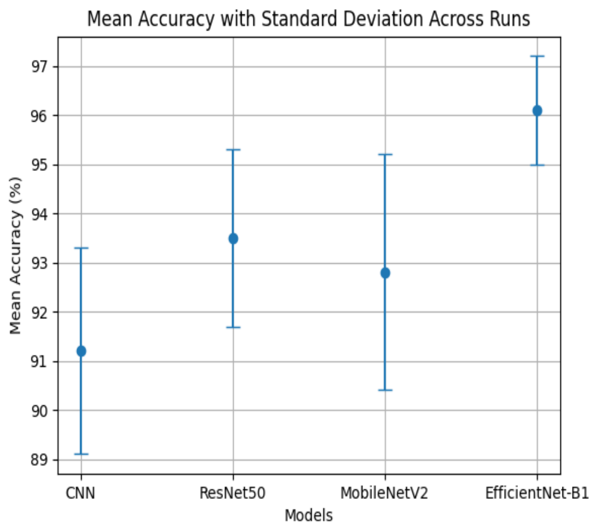

Figure 11 Mean Accuracy with Standard Deviation Across Multiple Runs for Different Models

This statistical validation strengthens the experimental findings and confirms that the observed improvements are consistent and reliable across repeated trials.

Conclusion

Figure 11 depicts The proposed EfficientNet-B1–based framework effectively detects both animal species and human presence in forest environments with high accuracy and reliability. By combining transfer learning with fine-tuning, the model demonstrates strong classification performance across varying lighting, weather, and motion conditions. It outperforms conventional architectures such as ResNet and YOLO, achieving superior precision and inference speed while maintaining computational efficiency. Through its advanced feature extraction capability and optimized scaling strategy, EfficientNet enables the system to distinguish between multiple species and identify potential human intrusions, contributing significantly to real-time forest surveillance and biodiversity protection. The integration of data preprocessing, augmentation, and EfficientNet's compound scaling ensures that the model performs robustly even with limited and diverse datasets.

REFERENCE

[1]. T. Lin, P. Goyal, R. B. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017 IEEE International Conference on Computer Vision (ICCV), IEEE, Venice, Italy, 2017, pp. 2999–3007.

[2]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S.-E. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," 14th European Conference on Computer Vision (ECCV), Springer, Amsterdam, The Netherlands, 2016, pp. 21–37.

[3]. J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, USA, 2017, pp. 6517–6525.

[4]. Y. Sakai, H. Lu, J. K. Tan, and H. Kim, "Recognition of surrounding environment from electric wheelchair videos based on modified YOLOv2," Future Generation Computer Systems, vol. 92, Elsevier, 2019, pp. 157–161.

[5]. S. Ren, K. He, R. B. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, IEEE, 2017, pp. 1137–1149.

[6]. J. Dai, Y. Li, K. He, and J. Sun, "R-FCN: Object detection via region-based fully convolutional networks," Advances in Neural Information Processing Systems (NeurIPS), vol. 29, 2016, pp. 379–387.

[7]. T. Lin, P. Dollár, R. B. Girshick, K. He, B. Hariharan, and S. J. Belongie, "Feature pyramid networks for object detection," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, USA, 2017, pp. 936–944.

[8]. S. U. Sharma and D. Shah, "Design and development of animal detection algorithm using image processing," Ph.D. thesis, Electronics and Communication Engineering, Gujarat Technological University, India, 2017.

[9]. A. Strandburg-Peshkin and F. H. Jensen, "Challenges and solutions for studying collective animal behavior in the wild," Philosophical Transactions of the Royal Society B, vol. 373, no. 1746, 2018, Article no. 20170005.

[10]. W. Liu, Z. Wang, N. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," Neurocomputing, vol. 234, Elsevier, 2017, pp. 11–26

[11]. X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3D object detection network for autonomous driving," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017, pp. 1907–1915.

[12]. B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2018, pp. 8697–8710.

[13]. B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million

image database for scene recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 6, IEEE, 2018, pp. 1452–1464.

[14]. M. C. Stoddard and D. Osorio, "Animal coloration patterns: Linking spatial vision to quantitative analysis," The American Naturalist, vol. 193, no. 2, 2019, pp. 164–186.

[15]. E. Karpestam, S. Merilaita, and A. Forsman, "Size variability effects on visual detection are influenced by colour pattern and perceived size," Animal Behaviour, vol. 143, 2018, pp. 131–138.