International Research Journal of Multidisciplinary Scope (IRJMS), 2024; 5(3):285-302

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2024.v05i03.0697

Safeguarding Humans from Attacks Using AI-Enabled (DQN) Wild Animal Identification System

Govindaprabhu GB[1]*, Sumathi M[2]

[1]Madurai Kamaraj University (MKU), Madurai, Tamil Nadu, India, [2]PG and Research Department of Computer Science, Sri Meenakshi Govt. Arts College for Women, Madurai, Tamil Nadu, India. *Corresponding Author's Email: prabhupri.pp@gmail.com

Abstract

Without advanced artificial intelligence (AI) technologies, monitoring and identifying wildlife has become increasingly difficult. To examine AI-driven methodologies for wild animal identification, this work uses a diverse dataset of annotated images with human, domestic and wild animal annotations. Convolutional Neural Networks (CNNs), AlexNet, and Deep Q-Learning (DQN) models are developed and compared by combining sophisticated preprocessing techniques such as dynamic color space conversion and day-night image translation. The models are evaluated on accuracy, precision, recall, F1-score, and mean percent error (MPE) loss metrics for classifying diverse species. The DQN model achieves the best performance with 79.5% accuracy, 0.78 precision, 0.84 F1-score, and 0.24 MPE loss. These findings demonstrate AI's potential to support conservation efforts by enabling accurate and automated wildlife monitoring. The comparative assessment of different models and factors influencing performance provides methodological insights to guide future research toward robust and generalizable AI solutions for biodiversity and habitat management.

Keywords: Artificial Intelligence, Conservation, Convolution Neural Network, Deep Q-learning, Wild Animal Identification, Wildlife Monitoring.

Introduction

Human populations expand into wild habitats, posing inherent risks and conflicts with wildlife. According to the Wildlife Conservation Society, over 500 species are in danger from retaliatory killings linked to human-wildlife conflict globally (1). These conflicts led to the killing of 550 elephants in Myanmar (2) and over 220 crocodile deaths per year in Indonesia (3). Furthermore, these encounters result in the loss of human life. For instance, recent reports document over 500 deaths per year in India resulting from elephant rampages across villages (4). Machine learning and artificial intelligence (AI) have made wildlife monitoring processes more automated and improved. AI-based approaches, such as deep learning, can identify images and audio recordings accurately. A large dataset of annotated wildlife images has enabled scientists to develop models that classify species with high accuracy and efficiency. AI has a key role to play in wildlife conservation and management. Artificial intelligence is used to collect, process, and analyze data about wildlife behavior, habitat monitoring, species identification, and more. By combining AI with wildlife, researchers and conservationists can gain valuable insights into various species and habitats to make informed decisions about their preservation and protection. This work explores a wild animal identification system using artificial intelligence. Convolutional Neural Networks (CNN), AlexNet, and Deep Q-Learning (DQN) are examined for identifying wild animals in diverse environmental conditions. Wild animal images are preprocessed using data augmentation, color space conversion, and day-to-night image conversion. There are three goals to this work: evaluating the performance of deep learning models for identifying wild animals; identifying factors affecting model performance, such as dataset quality, preprocessing methods, and model architecture; and discussing pragmatic importance of this work. As human populations expand into wild habitats, conflict between humans and wildlife is likely to increase, posing a threat to both. Mitigating these conflicts and supporting conservation efforts require effective monitoring and identification of wild animals. It can, however, be labor-intensive, time-consuming, and error-

(Received 02nd March 2024; Accepted 09th July 2024; Published 30th July 2024)

prone to monitor wildlife using traditional methods. To mitigate risks and safeguard human lives, it is important to employ more efficient and accurate approaches to mitigate human-wildlife conflicts. Advanced technologies with the capability of automating wild animal identification and enhancing proactive human protections are a pressing need in this context. A critical gap exists in the development and evaluation of AI-enabled systems specifically designed to address human safety concerns regarding wild animal identification.

The primary goal of this research is to create and test an AI-based system that can precisely identify and track wild animals in various natural environments using deep learning methods. More specifically, the aims of this work are as follows. AI-enabled wild animal identification system applications for wildlife conservation, habitat management, biodiversity preservation, and human-wildlife conflict mitigation. To build a diverse dataset of annotated images including human, domestic, and wild animal annotations for training and evaluating comprehensive models. This aims to assess the impact of preprocessing techniques, such as color space conversion, day-to-night image translation, and data augmentation, on animal identification performance. To provide methodological insights into dataset quality, preprocessing methods, and model architecture that influence deep learning model performance. This research investigates the performance of various DL models in identifying animal species from images, including CNN's, Alex-Net, and DQN.

There are several significant benefits to developing and implementing a robust wild animal identification system. A primary benefit is the mitigation of risks associated with conflicts with wildlife, and then it identifying wild animals contributes to conservation efforts by providing valuable data about wildlife populations and distributions. Intelligent identification systems are capable of processing data from diverse sources, including satellite imagery, camera traps, and acoustics, in an efficient and scalable manner. As well as educating the public about wildlife conservation, wild animal identification systems can promote coexistence with nature. By using AI-driven identification systems, costs can be reduced compared to traditional monitoring methods, since fewer human resources are required and existing infrastructure, such as camera traps, can be more effectively utilized. This research contributes to this work, in this study assessing how deep learning models perform in identifying wild animals to develop AI-based wildlife monitoring solutions. This research contributes to enhancing conservation efforts by developing automated species identification tools. This work offers methodological insights into future AI-based conservation research through the evaluation of model performance and influencing factors. Our research fosters interdisciplinary collaboration among biologists, ecologists, and economists. This work examines how wildlife monitoring methodologies have evolved historically and how they are used today. To identify wild animals, the method employs a comprehensive data acquisition strategy, sophisticated preprocessing pipelines, and tailored deep learning models, as discussed in section 3. Section 4, Results and Discussion, discusses the efficacy and nuances of deep learning models. In section 5, AI holds the potential to transform wildlife conservation while balancing ethical and societal concerns.

This research introduces an innovative DQNAI-enabled system for wild animal identification that excels in dynamic natural environments. The system integrates advanced preprocessing techniques with machine learning models, ensuring robust performance under diverse conditions. Trained on a combination of real and synthetically altered images, it maintains high accuracy even in challenging scenarios like low-light settings. This represents a significant advance over traditional methods, offering practical benefits for both human safety and wildlife management. By providing real-time animal detection and identification, the system enables early threat warnings and efficient population monitoring, supporting conservation efforts and mitigating human-wildlife conflicts. Its integration with various monitoring technologies makes it a valuable tool for biodiversity preservation and community protection, bridging the gap between academic research and real-world ecological and safety applications.

The following contents are described about the existing research work of wild animal identification. It finds factors that influence outdoor animal recognition and counting. First, it examines the types, uses, and mounting locations of sensors. Then it summarizes several studies related to livestock and wild animals (5). Various factors, including the px/m ratio, animal color, animal behavior, artificial objects, and the px/m ratio of the images, influence algorithm accuracy. Some traits are shared by wild and livestock animals. DeneseNet169, InceptionResNetV2, ResNetV2-50, and Xception were evaluated in (6). This Model performance is measured by F1 score and precision. The Inception and ResNetV2 outperformed the others by 99%. To avoid human interaction in forest-prone areas, the research developed an elephant intrusion monitoring system. A vision-based camera compares images with the template's images to identify animals. Elephant ivory, trunks, ears, and trunk patterns identify the animals. A LabVIEW algorithm is used to process images in this model (7). The research work used the Classifier with TESPAR data using Time and Spectrum features and Teager-MFCC spatial coefficients. In spite of the limited number of training data (sound recordings), TESPAR S-matrices were found to be effective at discriminating between 10 species in (domestic and wild animals, human voice recordings) (8). Using an identification device, the work proposed emitting a frequency that affects the nervous system of animals, which will cause them to leave that site. Biology and acoustics are combined in this cross-disciplinary science. Animals or humans use sound to communicate, so every sound or call is important (9).

In a paper proposed a system using a Wi-Fi microcontroller and the Internet of Things to detect wild animals near agricultural farms. Forest officers receive information from the transmitter using Energia IDE. Each corner also has RF transceivers, laser detectors, laser diodes, and buzzers. The Wi-Fi module sends a message on infringement. The proposed system is tested on an animal database. A Python server alerts the forest officer. This system can make high-casualty areas safer for humans and wildlife (10).

YOLOv5 is proposed to detect four types of farming intrusion animals. Cross-stage partial networks (CSP) serve as the backbone of YOLOv5. Input images are processed by this network to extract beneficial characteristics. In approximately 94% of cases, this method detected animal intrusions very effectively (11). All state-of-the-art criteria are met by these models. Activity engagement and duration are influenced by bear characteristics and external factors, according to the research (12). Furthermore, the methodology provides insight into the relationship between climate variables (temperature and precipitation) human activity (hunting), and animal age, gender, and activity engagement.

A system proposed YOLOv8 with deep learning algorithm is used here to detect four distinct categories: Lions, Tigers, Leopards, and Bears. Using documentaries, YouTube videos, and Kaggle datasets. A total of 1619 images in four categories are annotated. In addition to YOLOv8m, YOLOv8l, and YOLOv8x, three other YOLOv8 architectures were trained (13). The model proposed to improve by augmenting the dataset images. An extra-large model trained at 20 frames per second had a map of 94.3%. In-the-wild image capture with unconstrained conditions. Part-based convolutional networks (PCNs) represent discriminative part-level features. To alleviate the effects of the small amount of yak image data, random erasure and region-visibility prediction (RERP) are proposed as an auxiliary learning task. The proposed method with the SEResNet50 backbone achieves 97.57% Rank-1 accuracy and 76.30% mAP, respectively, compared to existing methods. The proposed method with the ViT backbone gets the best results when generalized to different views (14). The work used feature stride shortening, anchor size optimization, and hard negative class to overcome practical issues (15). A work examined 23,748 images from 14 UAS campaigns for the presence of kiang. Small animal detection performance was improved by researchers (16) by shortening feature strides and optimizing anchor sizes, respectively, and hard negatives (17) significantly reduced false positives, improving the F1 score from 0.44 to 0.86. A majority of the existing reviews focus on CNN's applications in different scenarios instead of addressing CNN in a more general sense, and some new ideas have not been discussed. The objective in this review is to provide some novel ideas and perspectives in this rapidly expanding area (18). Using 284,000 pre-segmented Urdu handwriting characters from 200 males and 200 females, two gender classification models are proposed, trained, and tested. Compared to existing deep learning gender classification models, the proposed models achieved state-of-the-art performance. Overall accuracy of the Alex-Net model was 99.14%, while that of the LeNet-5 model was 98.55% (19). Aviation is explored in this paper using Deep Reinforcement Learning. A Double Deep Q-Learning agent will be trained to control the plane's attitude control. The QPlane toolkit will be used for this and both simulators used (20).

The report provides a comprehensive overview of the HWC issue, based on relevant case studies and key lessons learned. With the exception of humans and elephants, this is a comprehensive review of wild mammal-human conflict written in the past decade (21). The spatial pattern of conflict is essential to understanding the dynamics of human-wildlife conflict. Human conflict with Asian elephants Elephas Maximus has increased in the Rajaji-Corbett landscape of Uttarakhand, India, where elephant habitat has been converted to agricultural land. Binomial Generalized Linear Models (GLMs) were used to analyze the predictors of household-level human-elephant conflicts (HECs) near protected areas using 266 semi-structured questionnaires (22). By using Pearson's bivariate chi-square test and binary logistic regression analysis, it intends to conduct an in-depth analysis of how attitudes toward HEC are influenced by location, demographics, and socio-economics in BTR and its neighboring areas. EDER includes the Eastern Doors Elephant Reserve (BTR) (23). Ecosystems depend on Asian elephants. Managing conflict between humans and elephants requires understanding this species' potential distribution area. The Global Biodiversity Information Facility (GBIF) data was used to simulate the potential distribution area of Asian elephants across South and Southeast Asia using maximum entropy (MaxEnt) (24, 25).

Many limitations exist in existing systems for identifying and monitoring wild animals. There are challenges relating to accuracy, such as poor image quality or low light, which can lead to misclassification. The recognition of diverse fauna is also challenging in many systems because they are species-specific. Computing resources are another barrier, making deployment in remote areas difficult. Developing and deploying systems is complicated by data bias and privacy concerns. Most conservation organizations, especially those in resource constrained areas, may not be able to afford sophisticated monitoring systems. Enhancing wild animal identification systems to support wildlife conservation requires addressing these limitations.

Methodology

The proposed work introduces novel advancements for wild animal identification systems. The first step is to improve preprocessing by incorporating innovative techniques like dynamic color space conversion, tailored to diverse environments, and advanced algorithms to ensure robust performance under varying lighting conditions. In addition, it integrates state-of-the-art deep learning models, including Convolutional Neural Networks (CNNs), AlexNet, and Deep Q-Learning. The models leverage hierarchical features and reinforcement learning techniques to identify and track wildlife with unprecedented accuracy and efficiency. To provide a comprehensive assessment of model performance, this work introduces a novel evaluation metric, Mean Percentage Error Loss (MPE), which surpasses traditional metrics. This work represents a significant advancement in the field of wild animal identification by combining these advancements. For wildlife monitoring and conservation, it offers unparalleled accuracy, adaptability, and reliability.

To enhance the model's robustness and generalization of the models across different scenarios and image characteristics the following methods are employed.

1. Lighting Conditions: The day-to-night conversion technique is employed to deal with diverse lighting scenarios.
2. Background Diversity: Images from various settings were included, and augmentation techniques were used to further diversify backgrounds.
3. Animal Pose and Orientation: To represent a wide range of poses and animations, extensive data augmentation methods are employed.
4. Image Quality: The training set included images of varying quality, and preprocessing steps were applied to enhance images where necessary.

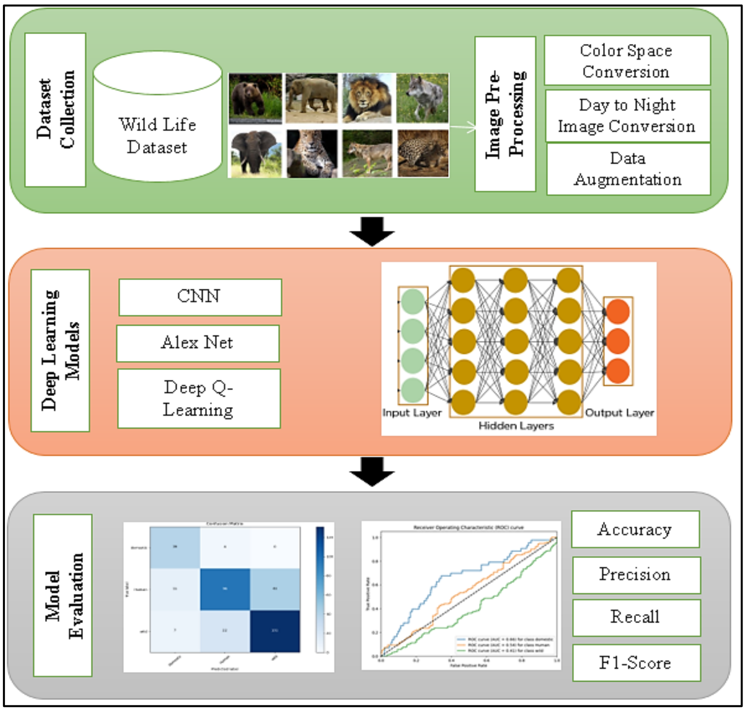

Figure 1: Flow of the Proposed Work

Dataset

Kaggle's dataset consists of training and testing sets, containing RGB images captured during the day. There are domestic animals, wild animals, and humans in these images. With this dataset, a system for detecting and alerting to wild animals' presence during the day is likely to be developed. It could be used for wildlife conservation, monitoring, safety, and conflict mitigation.

Dataset of this work were collected using high-resolution cameras (trap cameras, handheld cameras), which captured diverse images of wild animals in their natural habitat. A motion-activated HD camera trap is used to record motions. While capturing images during the day and night, it is equipped with infrared sensors for the significant feature set. Using professional cameras, wildlife photographers get close-up shots of various animals in different poses.

Preprocessing

Data pre-processing is an essential step in data analysis and machine learning, including the identification of wild animals. This preparation stage enhances raw data quality, reduces noise, and ensures compatibility with subsequent processing steps. A range of pre-processing techniques are used in wild animal identification to optimize their effectiveness. In the wild, lighting conditions vary from day to night, so color space conversion and data augmentation strategies are frequently used to standardize image representations.

Color Space Conversion

Using different color models or color spaces to represent colors in an image is called color space conversion. Based on mathematical principles, each color space defines colors differently. LAB (Lightness, A, and B color opponent dimensions) is the most commonly used color space in image processing and computer vision. Depending on the color space, animals may have distinct color patterns and characteristics. By converting images into different color spaces, it can enhance computer vision algorithms' ability to discriminate between different types of animals. Figure 2 describes the result of Color Space Conversion.

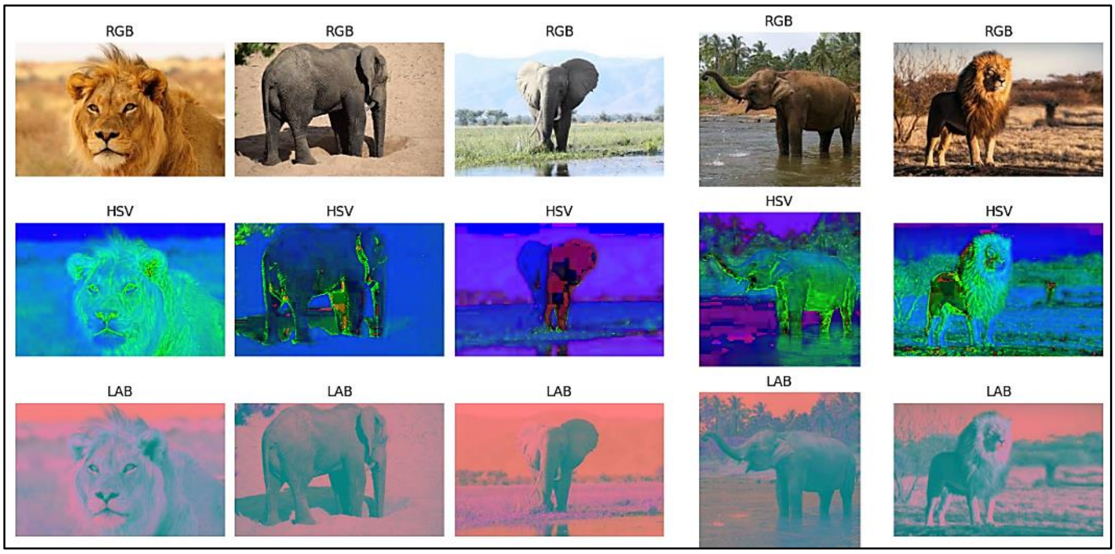

Figure 2: Color Space Conversion

Day to Night Image Conversion

For wild animal detection systems, day-to-night image conversion is essential to adapt to low-light environments, augment training data, simulate nighttime conditions for testing and evaluation, and enhance safety. Detecting and monitoring wild animals during both day and night can be achieved using these systems by converting daytime images into nighttime equivalents. Nighttime image collection in the wild can be expensive and challenging. Converting daytime images into nighttime equivalents enhances the training dataset, allowing the algorithm to learn from more diverse examples. The augmentation improves the robustness and generalization of the detection model, allowing it to recognize animals in different lighting conditions.

Algorithm 1: Day Night Conversion

The algorithm (Table 1) converts daytime images to simulated nighttime images through four steps. First, each pixel's brightness is decreased by a fixed amount. The second step is to add a blue tint by increasing the blue channel while leaving red and green unchanged. According to a supplied kernel size parameter, it applies Gaussian blur to smooth over the image. Lastly, it returns the processed image with blue tint, Gaussian blur, and decreased brightness. Through strategic adjustments to brightness, color, and sharpness, the algorithm simulates nighttime scenes from daytime images. Figure 3 shows the Night image converted results.

Table 1: Algorithm 1: Day Night Conversion

Input: Day Time Image (I_{Day})

Output: Night Time Image (I_{Night})

Algorithm: Day_Night_Conversion

1. Let I_{Day} represent the daytime image loaded from the given image path
2. For each pixel P, in I_{Day}, adjust its brightness using the following formula
3. $V_P = clip(V_P + brightness\_factor, 0{,}254)$. where, V_P is the value (brightness) component of pixel p, and brightness_factor is a negative value indicating the amount of brightness adjustment
4. Apply a blue tint to the image I_{Day}. Let I_{Tint} be the resulting image after applying the blue tint. The blue tint can be applied by adding a constant blue color matrix T to each pixel in I_{Day}. $I_{tint} = I_{day} + T$, Where T= [-10 0 10].
5. Apply Gaussian blur to the image I_{Tint} to add a softness effect, $I_{Tint} = GaussianBlur(I_{Tint}, Kernel\_size)$, where kernel_size is the size of the Gaussian kernel used for blurring.
6. Return the Processed Image I_{Night}

End Algorithm

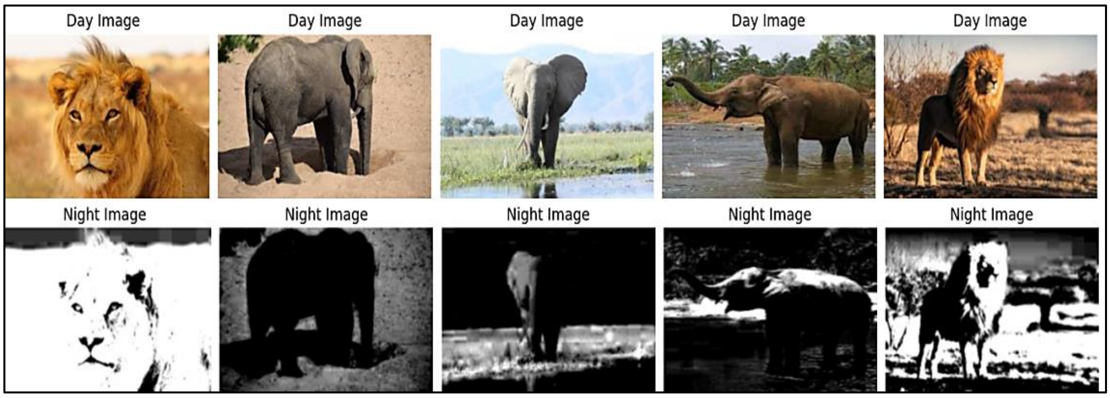

Figure 3: Day to Night Converted Image

Data Augmentation

Image augmentation (14) involves applying various transformations to existing images to artificially increase the diversity of a dataset. The collection of a large and diverse dataset of wild animal images can be time-consuming and challenging. With image augmentation, it can generate more training examples by rotating, scaling, flipping, cropping, and translating the existing images. This increases the size of the dataset, allowing the detection model to be trained on more data. There are many poses, orientations, and lighting conditions that animals can appear in in the wild. The augmentation of images during training makes the detection model more robust to these variations. Models can generalize to unseen examples in real-world scenarios by training on augmented data.

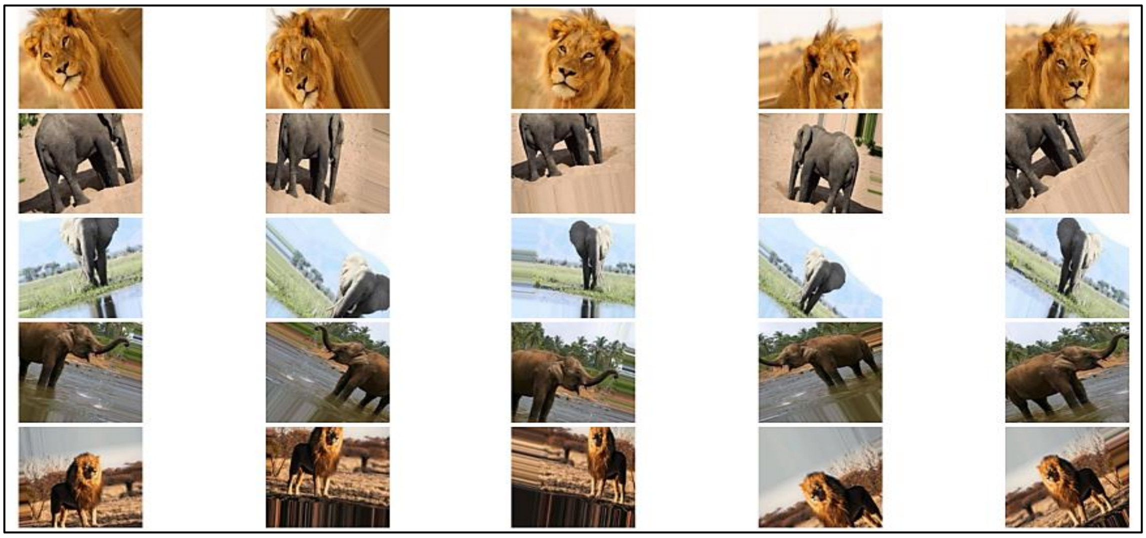

Figure 4: Data Augmented Images

Figure 4 shows the augmented data images. Class imbalance issues can occur in wildlife datasets when certain animal classes dominate. Using image augmentation, synthetic images can be generated augmentation, synthetic images can be generated for underrepresented classes, ensuring that all classes receive sufficient training data.

Deep Learning Models

A subset of machine learning algorithms, deep learning models employ multiple layers of interconnected neurons (ANNs) to learn hierarchical representations of data. Computer vision, natural language processing, and speech recognition have all benefited from using these models because they extract complex patterns and features from raw data. Deep learning models are essential for developing accurate and efficient detection systems for wild animals.

(a) Training Set: These images are mostly used for training (70%) with domestic animals, wild animals, and humans. The model to learn and understand the characteristics of different types of animals and humans, these images are required.

(b) Testing Set: To evaluate the trained model, 30% of the set is tested. The model has not seen these images in the training phase, but they are similar to those in the training set. This allows researchers and developers to assess the model's ability to predict new data.

Data Annotation Process is done with training the AI-enabled identification system prepared in the following manner. At first, the Experts manually reviewed and annotated each image by manual labeling. To ensure accurate and detailed annotations, each animal was identified and labeled according to its species. Objects were detected in the images by drawing bounding boxes around each animal. Using this technique, it could identify animals in different backgrounds.

Convolutional Neural Network

In deep learning, convolutional neural networks (CNN/ConvNet) (15) are a class of deep neural networks used to analyze visual imagery. Neural networks are typically associated with matrix multiplications, but ConvNet is not. The technique is called convolution. In mathematics, convolution is the process of modifying the shape of two functions by producing a third function.

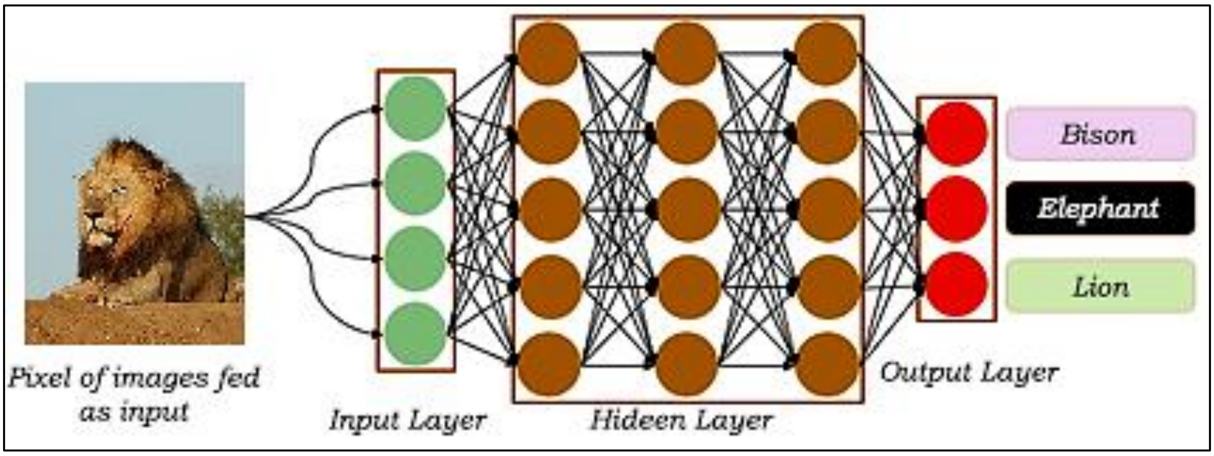

Figure 5: Convolutional Neural Network Architecture

Figure 5 shows the architecture of Convolutional Neural Network. In CNNs, convolutional layers apply filters (also known as kernels) to the input images. Edges, textures, and shapes can be extracted from the input images using these filters. During wild animal detection, the convolutional layers learn discriminative features that distinguish animals from background scenes. To down sample feature maps and reduce spatial dimensions, pooling layers, such as max pooling or average pooling, are often inserted after convolutional layers. By pooling the learned features, the models become more resilient to translations and distortions of the input images. By pooling layers, wild animal detection models can capture the most salient features while reducing computational complexity.

In a fully connected layer, the extracted features are flattened after several layers of convolutional and pooling operations. Feature extraction from input images is used to train these layers to learn complex patterns and relationships between features. An image containing a wild animal is classified as either containing a particular type of animal or belonging to the background class by fully connected layers. A non-linear activation function, such as ReLU (Rectified Linear Unit), is applied after each convolutional and fully connected layer to introduce non-linearity and enable the network to learn complex mappings. ReLU is commonly used because of its simplicity and effectiveness in preventing vanishing gradients. CNN architectures consist of one or more neurons as their output layer, depending on their specific function. The presence or absence of a particular animal is commonly classified as a sigmoid-activated neuron. The output layer for multiclass classification tasks (e.g., classifying multiple types of animals). Figure 6 shows the Confusion Matrix of CNN.

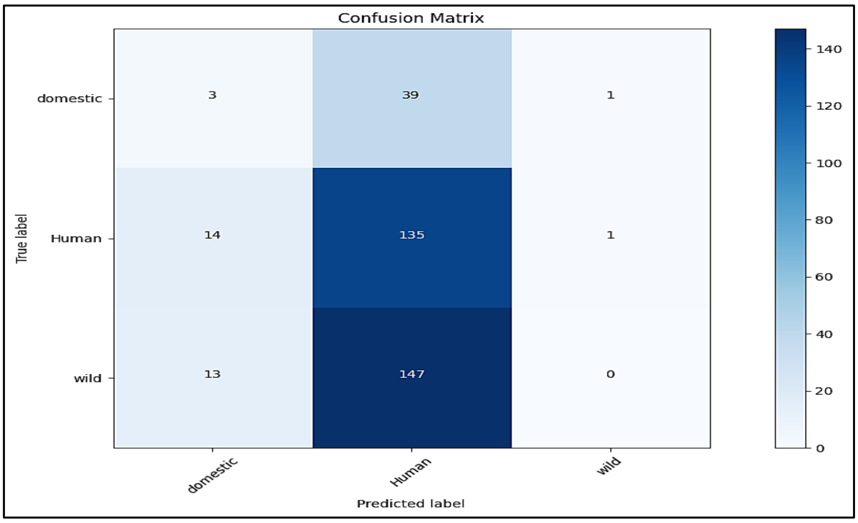

Figure 6: Confusion Matrix of CNN

Algorithm 2: CNN Model

Algorithm the CNN model (Table 2) has 4 convolutional layers interspersed with 4 max pooling layers, followed by 2 fully connected layers prior to output. ReLU activation is used, and the network is trained with categorical cross-entropy loss optimized using Adam.

Alex Net

AlexNet (16) has 5 convolution layers, 3 max-pooling layers, 2 normalized layers, 2 fully connected layers, and one SoftMax layer. An activation function called ReLU is added to each convolution layer. Because of the presence of fully connected layers, the input size is fixed because of the max-pooling function. The input size is usually stated as 224x224x3, but due to padding it ends up being 227x227x3. There are over 60 million parameters in AlexNet. Figure 7 shows the architecture of the Alex Net.

Table 2: Wild Animal CNN Algorithm

Input: Training dataset (X_train, y_train): A collection of labeled images of wild animals and background scenes. Testing dataset (X_test): Unlabeled images for evaluating the trained model.

Output: Predictions & Accuracy

Algorithm: Wild Animal CNN

1. Define an input layer of 150x150 pixels with three color channels corresponding to the input images with Training Set X_train, y_train and Testing Set X_Test.
2. Activate four convolutional layers with 32, 64, 128 and 128 filters, respectively.
3. After each convolutional layer, use MaxPooling layers to downsample the feature maps.
4. Finalize the convolution by flattening the output.
5. Activate ReLU and connect 512 units in a fully connected layer.
6. Regularize dropouts with a 0.5 rate.
7. Multi-class classification with three classes requires softmax activation at the output layer.
8. Measure the difference between predicted and actual distributions using categorical cross-entropy. and Choose Adam optimizer with default learning rate.
9. The accuracy metric should be monitored during training.
10. Use the data generator with a batch size of 32 to feed the training dataset into the model.
11. Using backpropagation and gradient descent optimization, train the model over 10 epochs.
12. Make predictions on the testing dataset using the trained model (not provided in the code snippet).
13. Detect wild animals accurately by calculating accuracy.

End Algorithm

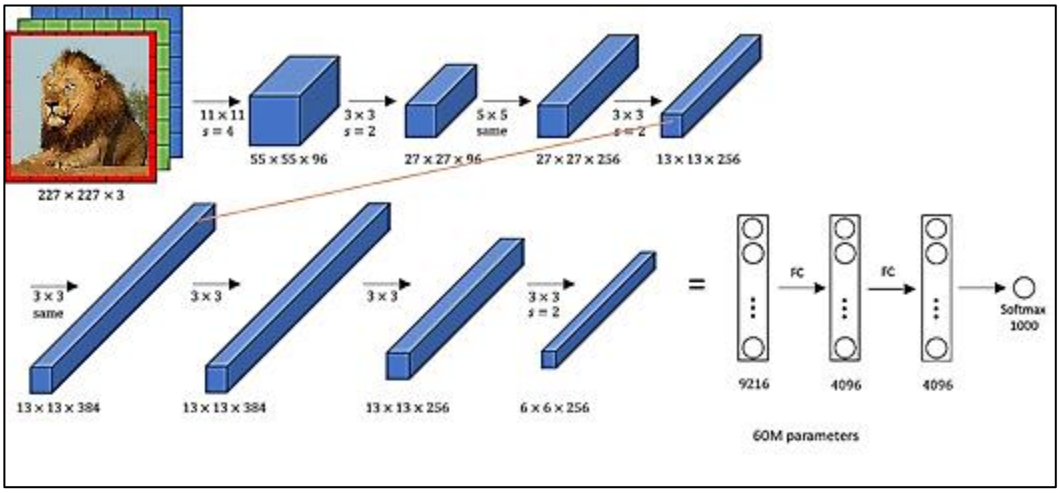

Figure 7: Architecture of the Alex Net

AlexNet uses images with dimensions of 227x227x3, where the three channels represent color information (red, green, blue). Wildlife habitats are featured in these images, along with a variety of animals and scenery. A Rectified Linear Unit (ReLU) activation function follows each convolutional layer. Convolutional layers can extract hierarchical features from input images, allowing the model to distinguish between types of animals based on fur texture, body shape, and distinctive markings. A max-pooling layer is interspersed between the convolutional layers. In max-pooling operations, the spatial dimensions of feature maps are reduced while the most salient features are retained. Using this method, the model focuses on the most informative aspects of the images and can be generalized across different wildlife scenes. After the convolutional layers, AlexNet incorporates local response normalization (LRN). Its robustness to variations in lighting, weather conditions, and camera angles is improved by these layers, which enhances its ability to detect subtle variations in animal appearances. Two fully connected layers follow the convolutional and pooling layers. A model can capture complex relationships between animal attributes and background elements by learning high-level representations of features extracted from input images. Fully connected layers are crucial for predicting the presence of wild animals. In AlexNet, a soft-max layer outputs the probability distribution over different classes of wild animals. The model was trained to recognize different types of animals by classification. With the soft-max activation function, the model calculates a probability score for each class of animal, indicating whether that animal is likely to appear in the input image. Figure 8 shows the Confusion matrix of the alex net model.

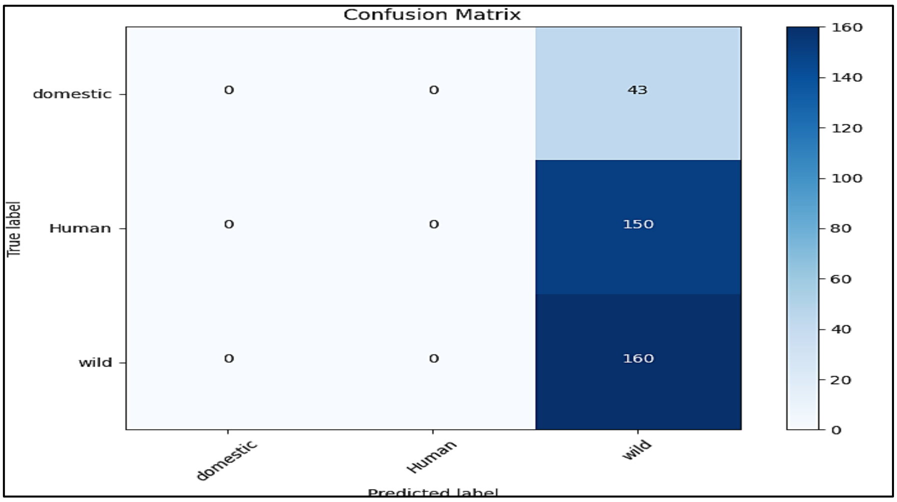

Figure 8: Confusion Matrix of Alex Net

Table 3: Algorithm for Wild Animal AlexNet

Input: Training dataset (X_train, y_train): A collection of labeled images of wild animals and background scenes. Testing dataset (X_test): Unlabeled images for evaluating the trained model.
Output: Predictions for the testing dataset
Algorithm: Wild Animal AlexNet

1. A convolutional layer (Conv) with ReLU activation functions is added to the input layer with dimensions 227x227x3 with Training and Testing Set.
2. To down sample feature maps, intersperse three Max Pool layers.
3. After the first and second convolutional layers, add two Local Response Normalization (LRN) layers (Normalization).
4. To learn high-level representations, add two fully connected layers (FC).
5. A softmax output layer should be added to produce probability distributions over different animal classes.
6. To minimize the defined loss function, feed the training dataset (X_train, y_train) into the Alex Net model and adjust model parameters by backpropagation and optimization. During each epoch, the network is traversed forward and backward.
7. Predict the testing dataset (X_test) using the trained model.
8. Analyze the model's performance metrics to determine its effectiveness.

End Algorithm

Algorithm 3: WildAnimal AlexNet Model
Algorithm 3 (Table 3) shows the Alexnet model. AlexNet contains 5 convolution layers, 3 max-pool layers, and 2 fully connected layers. ReLU activation is employed, and the model is optimized through SGD to minimize log loss.

Deep Q-Learning

The deep Q-learning algorithm (21) combines the power of deep learning with Q-learning principles. Image classification has been successfully applied to various tasks. Deep Q-learning uses deep neural networks to approximate the Q-function, which represents the expected reward for taking particular action at a given state. To train the network, past experiences are stored in a buffer and randomly sampled using an experience replay technique. Figure 9 shows the architecture of the Deep Q Learning.

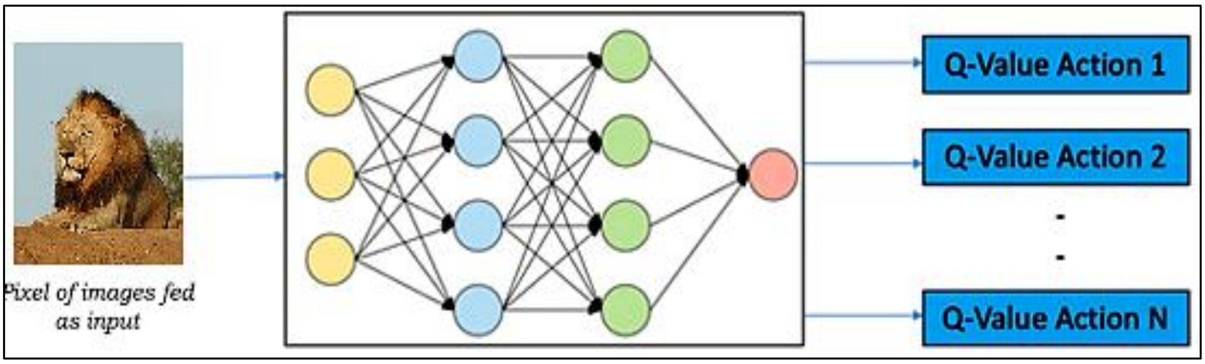

Figure 9: Architecture of the Deep Q-Learning

A camera placed in a wildlife habitat can capture images representing a state space in wild animal detection. Images provide information about the current environment, such as the presence or absence of animals, their locations, and potentially other contextual data, such as weather conditions. Agent actions are determined by the observed states. When wild animals are detected, the camera may be moved to different angles or positions, its settings adjusted (e.g., zoom level, exposure), or alerts triggered. Agents receive feedback based on their actions through the reward function. To detect and track wild animals accurately, a reward function may be designed to minimize false alarms and energy consumption. Positive rewards could be given for correctly detecting animals and negative rewards for incorrectly detecting animals. The Deep Q-Network (DQN) is DQL's core component, which is a deep neural network that outputs Q-values based on an image (state). In the current state, the Q-values represent the expected rewards for taking each action. By training, the DQN approximates the optimal action-value function, which specifies the expected cumulative reward for following a particular policy. Training and sample efficiency are typically improved through experience replay in the DQL. DQNs are trained by sampling mini-batches of experiences and storing them in a replay buffer. This prevents the network from overfitting to recent experiences by decorrelation. In addition, a target network may be used to stabilize training by providing target Q-values during the update process. By updating the target network periodically, the variance in the Q-value estimations is reduced and convergence is improved. Through Q-learning or deep Q-learning with experience replay, the agent interacts with the environment by selecting actions based on the current state. A wild animal detection policy is learned iteratively until an optimal policy is reached. Once trained, a DQL agent can autonomously detect and track wildlife in real time in wildlife monitoring systems. Detection accuracy, false alarm rate, and energy efficiency can be used to evaluate the agent's performance. Figure 10 shows the Confusion Matrix of Deep Q-Learning.

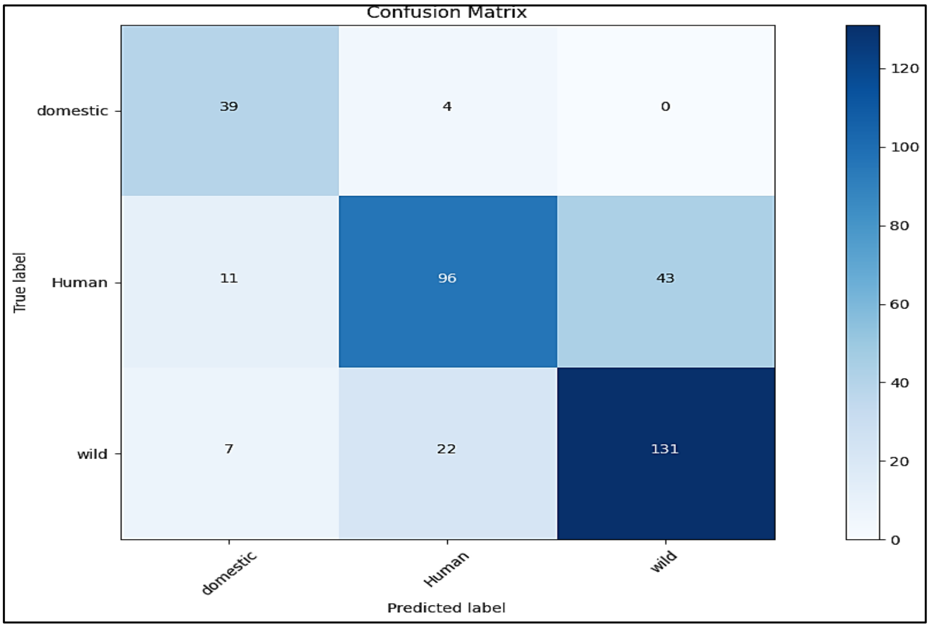

Figure 10: Confusion Matrix of Deep Q-Learning

Table 4: Algorithm for Deep Q-Network (DQN) for Image Classification

Input: Training dataset (X_train, y_train): A collection of labeled images for training the DQN, testing dataset (X_test, y_test): A collection of labeled images for evaluating the trained DQN, Neural network architecture, Replay buffer capacity, Discount factor (gamma), Exploration parameters (epsilon), Number of episodes

Output: Classification accuracy on the testing dataset

Algorithm: Deep Q-Network (DQN) for Image Classification

1. Create a random weighted DQN neural network.

2. Initialize the replay buffer with the capacity 'replay_buffer_capacity'.

3. For each training episode:

a. Initialize the environment (neural network parameters).

b. Select an image from the training dataset to initialize the initial state 's'.

c. While the episode is not finished:

i. Choose a random action 'a' with probability epsilon (exploration).

ii. Alternatively, select the action 'a' that maximizes the Q-value for the current state 's' (exploitation).

iii. Act 'a' (forward pass through the neural network) and observe next state 's_next' (output probabilities for each class).

iv. Compare predicted probabilities with one-hot encoded labels for the selected action.

v. Replay buffer the transition (s, a, loss, s_next).

vi. Take a mini-batch of replay buffer transitions.

vii. Determine the target Q-values using the Bellman equation:

Q_target(s, a) = loss

viii. Reduce the loss between predicted and target Q-values by updating the DQN parameters:

loss = MSE (Q_predicted(s, a), Q_target(s, a))

A gradient descent algorithm is used to update the DQN parameters.

ix. Change the current state 's' to the next state 's_

4. Using the training dataset, evaluate the trained DQN:

a. Using the testing dataset, for each image:

i. Apply the trained DQN to the image and obtain the predicted probabilities for every class. Assign the class with the highest probability as the predicted class.

b. Determine the classification accuracy by comparing the predicted classes to the ground truth.

5. Display the classification accuracy and trained DQN model on the testing data.

End Algorithm

Algorithm 4: DQN Model

Algorithm 4 (Table 4) shows the DQN Model. The DQN uses a ResNet-50 backbone for feature extraction. It has a convolutional encoder-decoder pathway to produce Q-values for each possible action. Experience replay memory size is 10,000 transitions, and target network update frequency is 100 steps.

This work prioritized data dependability and accuracy by sourcing a diverse image dataset from Kaggle, followed by particular manual labeling. Preprocessing techniques, including color space conversion and day-to-night image transformation, were applied to enhance image features. Data augmentation methods such as rotation, scaling, and flipping were employed to increase dataset diversity and improve model robustness. The Deep Q-Learning (DQN) system, effectiveness was evaluated using standard machine learning metrics (Accuracy, Precision, Recall, F1-Score) alongside a novel Mean Percentage Error Loss measure, ensuring a comprehensive assessment of the model's performance across various scenarios.

Results and Discussion

This work is implemented and tested with Intel Core-i7 2620M, 16 GB RAM, 1TB HDD with Windows 11 and Python. Data analysis, machine learning, and evaluation are performed using Python, which is used to implement the system. Data manipulation and numeric computing tools are provided by NumPy and Pandas. Scikit-Learn provides ML / DL algorithms. The Matplotlib and Seaborn libraries can be used to visualize data, features, and accuracy metrics.

Three deep learning models were evaluated on a set of metrics: CNN, AlexNet, and Deep Q-Learning (DQN). 1,346 images are included in the dataset from Kaggle, split into three categories: Humans, Domestic animals, and Wild animals. Among the wild animals are elephants, lions, bears, and others. Dogs, cats, and similar animals are domestic animals. Training and test sets are split 75-25%. In addition, unlabeled daytime images of wildlife were scraped from public repositories. The day-night conversion algorithm generated nighttime variants. The training images were augmented with 2,000. The models were trained in TensorFlow for 50/100 epochs with early stopping if validation loss doesn't decrease for 5 consecutive epochs. With a learning rate 0.001 and a batch size of 32, Adam optimizer was used. To balance the data, classes were weighted. Mixed precision models were trained using Keras and TensorFlow. Accuracy measures how often the model correctly predicts class labels. The error rate or misclassification rate represents the ratio of incorrect to accurate predictions. Figure 11 shows the accuracy and Error rate.

Figure 11: Accuracy and Error Rate

Precision measures how accurate the model is. This ratio measures the number of true positives made by the model (true positives plus false positives) versus the total number of positive predictions. In wild animal identification, precision refers to the proportion of correctly identified instances of a particular species among all predicted instances. The model is more precise when it identifies the target species with higher precision. The recall or sensitivity of a model measures its ability to identify instances of a given class. True positives and false negatives are calculated as the ratio of true positives to actual instances. Recall is the ratio of correctly identified instances to all actual instances of a species. Models with higher recall minimize false negatives by capturing more instances of the target species. A model's F1 score provides a balanced measure of precision and recall. Weighted average of precision and recall, ranging from 0 to 1. Precision and recall are balanced in high F1 scores. Wild animal identification models were evaluated using the F1 score to minimize false positives and false negatives. Figure 12 shows the precision, recall and F1-Score (26).

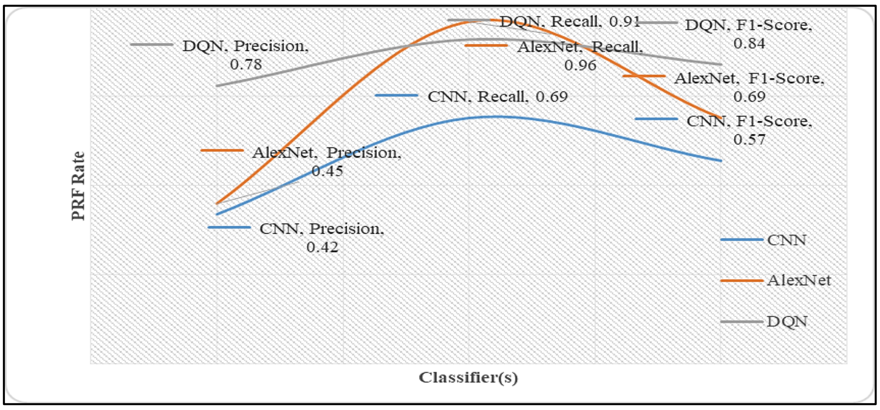

Figure 12: Precision, Recall & F1-Score

MFE loss is the average percentage difference between the predicted and actual values in a dataset. The mean of all absolute percentage errors in the dataset was calculated. When using AI to identify wild animals, MFE Loss may be applied when predicting continuous variables based on image data, such as animal size or age. The lower the MFE Loss, the better the performance of the regression model, as it reflects a smaller deviation between the predicted and actual values. Figure 13 shows the MFE Loss.

The results revealed significant variations in the performance of the models across different metrics. Notably, the DQN model exhibited the highest accuracy (79.5%), lowest error rate (20.5%), and superior precision (0.78) compared to CNN and AlexNet. Conversely, CNN and AlexNet demonstrated relatively lower accuracy (39.1% and 45.4%, respectively) and precision (0.42 and 0.45, respectively), although AlexNet displayed higher recall (0.96) compared to DQN and CNN.

The Receiver Operating Characteristic (ROC) curve is a graphical representation used to evaluate the performance of binary classification models. It illustrates the trade-off between the true positive rate (sensitivity) and the false positive rate (1 - specificity) across different thresholds used for classifying instances. Figure 14 shows the ROC Curve.

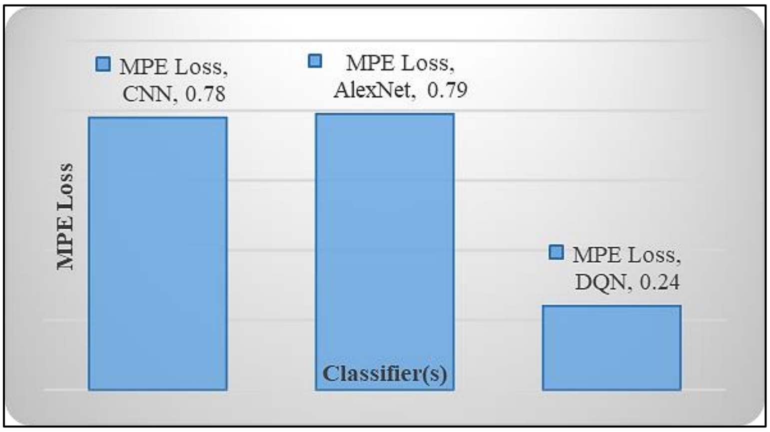

Figure 13: MPE Loss

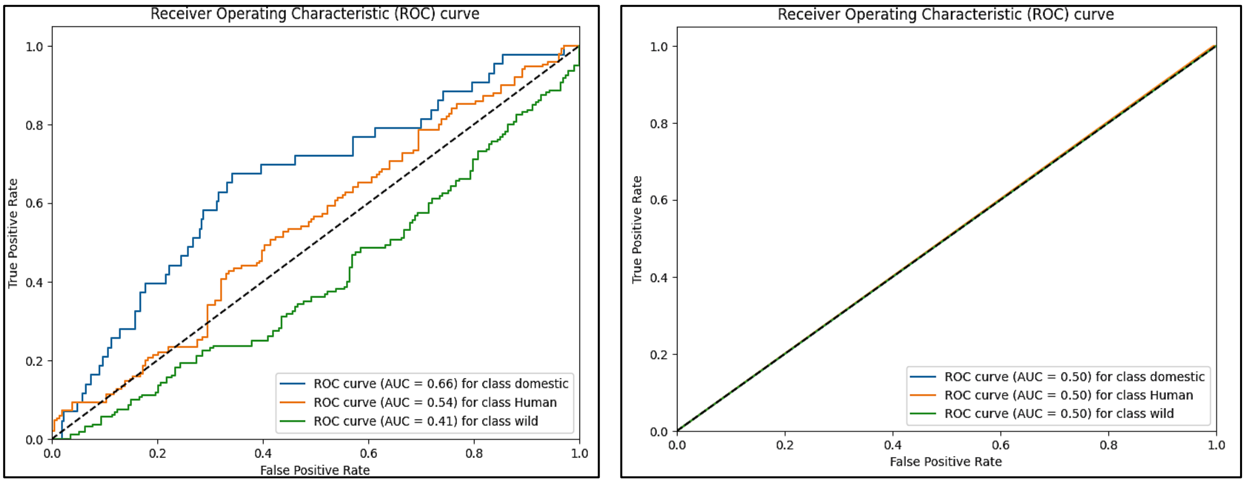

Figure 14: ROC Curve of CNN & Alex Net

Wild animal identification using AI, the ROC curve in Figure 14 and 15 provide valuable insights into the performance of binary classification models in distinguishing between different animal species. They help researchers assess the model's sensitivity to true positive identifications while controlling for false positive identifications, thereby guiding model selection and optimization efforts.

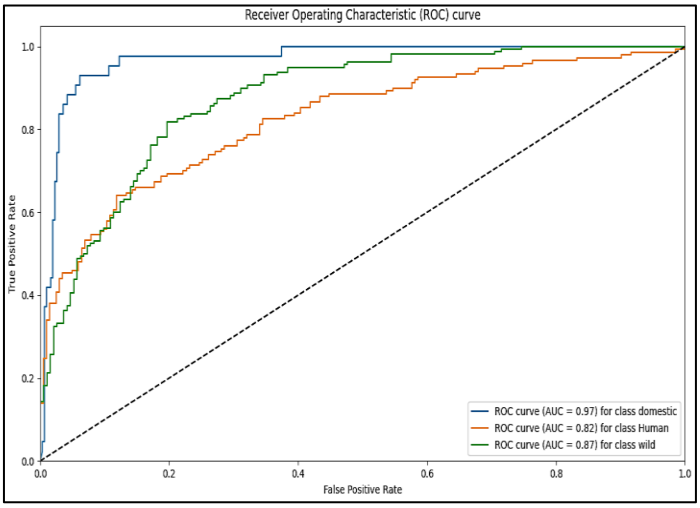

Figure 15: ROC of DQN

The system exhibits significant precision and steadfastness in distinguishing a broad range of animal species under diverse environmental scenarios, which is of substantial worth in the context of safeguarding wildlife. Its proficiency in effectively overseeing and monitoring animal communities is instrumental in facilitating well-timed actions concerning species at risk of death. Moreover, the system's capacity to reliably recognize animals in proximity to human settlements contributes to the mitigation of human-wildlife conflicts by offering early alerts and enabling defensive strategies. Furthermore, integrating this technology with camera traps, drones, and surveillance networks enhances real-time monitoring and decision-making capabilities. Ultimately, this research not only advances animal identification technology but also offers practical solutions for maintaining ecological balance and protecting communities, bridging the gap between technological innovation and real-world applications in wildlife management and public safety. There are several strengths to the choice of test environment and dataset for this work in terms of replicating real-life wild animal's interactions. A diverse dataset was used in the work, including images of humans, domestic pets, and wild animals. To distinguish the types of different diversity subjects the employed models are trained. This dataset likely contains a variety of natural backgrounds, lighting conditions, and animal poses/behaviors that real-life deployments would need to handle. It is difficult to capture low-light conditions which commonly occur while monitoring wildlife. Using data augmentation techniques, the training set was expanded to 2,000 images for exposing the models to a wider range of variations in the real world. The proposed Model to be determined if they are capable of identifying a variety of wildlife species such as elephants, lions, and bears.

Conclusion

This work presented an investigation into the effectiveness of various deep learning models for wild animal identification using AI. Leveraging a diverse dataset of wild animal images and employing sophisticated preprocessing techniques, including color space conversion, day-to-night image conversion, and data augmentation, model trained and evaluated three deep learning models: Convolutional Neural Network (CNN), AlexNet, and Deep Q-Learning (DQN). The results demonstrate significant variations in the performance of these models, with DQN outperforming CNN and AlexNet in terms of accuracy, error rate, precision, and overall predictive capability. The success of DQN underscores the potential of reinforcement learning-based approaches for tackling complex tasks such as wild animal identification, where sequential decision-making and environmental interactions play a crucial role.

While our findings offer promising insights into the feasibility of AI-based solutions for wildlife monitoring and conservation, several challenges and opportunities remain. Future research efforts should focus on addressing limitations such as dataset biases, model interpretability, and scalability to ensure the robustness and applicability of AI models across diverse ecosystems and species. Despite the promising results, several limitations should be acknowledged. These include the reliance on static image data, potential biases in dataset composition, and challenges associated with model interpretability and explainability. Future research directions may focus on addressing these limitations through the integration of dynamic sensor data (e.g., audio, video streams), collaborative data sharing initiatives, and the development of interpretable AI algorithms. Moreover, efforts should be made to evaluate the scalability and transferability of the proposed models to different geographic regions and ecological contexts.

Abbreviations

Nil.

Acknowledgments

No.

Author Contributions

The authors confirm sole responsibility for this work, includes conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Conflict of Interest

None.

Ethics Approval

Not applicable.

Funding

Nil.

References

1. Leimgruber P, Gagnon JB, Wemmer C, Kelly DS, Songer MA, Selig ER. Fragmentation of Asia's remaining wildlands: implications for Asian elephant conservation. Animal Conservation. 2003;6(4):347-359. doi:10.1017/S1367943003003421.
2. Bezuijen MR, Webb GJ, Hartoyo, Samedi, Congdon BC. Occurrence and conservation needs of crocodiles in Kalimantan, Indonesia. Crocodile Specialist Group Newsletter. 2010.
3. Sampson C, McEvoy J, Oo ZM, Chit AM, Chan AN. New elephant crisis in Asia: Early warning signs from Myanmar. PLOS ONE. 2018; 13(3):1-10. doi: 10.1371/journal.pone.0194113.
4. Nad C, Roy R, Roy TB. Human elephant conflict in changing land-use land-cover scenario in and adjoining region of Buxa tiger reserve, India. Environmental Challenges. 2022 Apr 1;7:100384. https://doi.org/10.1016/j.envc.2021.100384
5. Evstatiev B, Kalmukov Y. Analysis of the Key Factors Influencing the Outdoor Animal Recognition and Counting Accuracy. 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), Ruse, Bulgaria. 2022;8(1):1-9. doi: 10.1109/EEAE53789.2022.9831410.
6. Thangaraj R, Rajendar S, Sanjith M, Sasikumar S, Chandhru L. Automated Recognition of Wild Animal Species in Camera Trap Images Using Deep Learning Models. In2023 Third International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT). IEEE. 2023 Jan 5:pp. 1-5. doi: 10.1109/ICAECT57570.2023.10117922.
7. Ram AVP, Sarath Prakash A, Ahamed AI, Anirudh K, Arvindh MM, Nithyavathy N. A Self Induced Warning System for Wild Animal Trespassing Using Machine Vision System. 2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW), India. 2018;13(2):349-353. doi: 10.1109/I2C2SW45816.2018.8997520.
8. Pop GP, Vaida MF. Animal Species Identification Using TESPAR S-Matrices based Features. In2019 E-Health and Bioengineering Conference (EHB). IEEE. 2019 Nov 21:pp. 1-4. doi: 10.1109/EHB47216.2019.8969960.
9. Shetgaonkar A, Vinayak NSP. Smart animal repeller. 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), India. 2017;6(1):3303-3308. doi: 10.1109/ICECDS.2017.8390070.
10. Begum HM, Janeera DA, Kumar AG. Internet of Things based Wild Animal Infringement Identification, Diversion and Alert System. International Conference on Inventive Computation Technologies (ICICT), India. 2020; 20(1):801-805. doi:10.1109/ICICT48043.2020.9112433.
11. Mamat N, Othman MF, Yakub F. Animal Intrusion Detection in Farming Area using YOLOv5 Approach. 22nd International Conference on Control, Automation and Systems (ICCAS), Republic of Korea. 2022; 22(1):1-5. doi: 10.23919/ICCAS55662.2022.10003780.
12. Calastri C, Giergiczny M, Zedrosser A, Hess S. Modelling activity patterns of wild animals - An application of the multiple discrete-continuous extreme value (MDCEV) model. Journal of Choice Modelling. 2023; 47(2):1-11. doi: 10.1016/j.jocm.2023.100415.
13. Dave B, Mori M, Bathani A, Goel P. Wild Animal Detection using YOLOv8. Procedia in Computer Science. 2023; 230(1):100-111. doi: 10.1016/j.procs.2023.12.065.
14. Li L, Zhang T, Cuo D, Zhao Q, Zhou L, Jiancuo S. Automatic identification of individual yaks in in-the-wild images using part-based convolutional networks with self-supervised learning. Expert System Application. 2023; 216(1):1-10. doi: 10.1016/j.eswa.2022.119431.
15. Peng J, Wang D, Liao X, Shao Q, Sun Z, Yue H, Ye H. Wild animal survey using UAS imagery and deep learning: modified Faster R-CNN for kiang detection in Tibetan Plateau. ISPRS Journal of Photogrammetry and Remote Sensing. 2020; 169(2):364-376. doi: 10.1016/j.isprsjprs.2020.08.026.

16. Safibullaevna BS, Shukurovich MA, K kizi JM. Processing Color Images, Brightness and Color Conversion. In: 2019 International Conference on Information Science and Communications Technologies (ICISCT), Uzbekistan. 2019; doi: 10.1109/ICISCT47635.2019.9012040.
17. Rajasekaran Thangaraj, Charly Jerome J, Sanjith M, Rithick Saran K, Sudev Sasikumar, Vijayakumar S. Automatic Detection and Classification of Wild Animal Species Using YOLO Models,3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India. 2023; 3(1):315-322. doi: 10.1109/ICPCSN58827.2023.00058.
18. Li Z, Liu F, Yang W, Peng S, Zhou J. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning System. 2022; 32(12):6999-7019. doi: 10.1109/TNNLS.2021.3084827.
19. Nabi ST, Singh P, Kumar M. Gender Classification from Offline Handwriting Images in Urdu Script: LeNet-5 and Alex-Net. Third International Conference on Applied Artificial Intelligence (ICAPAI), Halden, Norway. 2023; (1-6):5-9.doi: 10.1109/ICAPAI58366.2023.10194140.
20. Richter DJ, Calix RA. Using Double Deep Q-Learning to learn Attitude Control of Fixed-Wing Aircraft. 16th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), Dijon, France. 2022; 16(1): 646-651. doi: 10.1109/SITIS57111.2022.00102.
21. Distefano E. Human-wildlife conflict worldwide: collection of case studies, analysis of management strategies and good practices. Food and Agricultural Organization of the United Nations (FAO), Sustainable Agriculture and Rural Development Initiative (SARDI), Rome, Italy. 2005.
22. Megha Rani, Debashish Panda, Maximilian L. Allen b, Puneet Pandey, Randeep Singh. Assessment and prediction of human-elephant conflict hotspots in the human-dominated area of Rajaji-Corbett landscape, Uttarakhand, India. Journal for Nature Conservation. 2024; 79(1):22-29. https://doi.org/10.1016/j.jnc.2024.126601.
23. Chiranjib Nad, Tamal Basu-Roy. Association between people's attitudes towards human-elephant conflict and their locational, demographic, and socio-economic characteristics in Buxa Tiger Reserve and its adjoining area, India. Regional Sustainability. 2024;5(1):1-10. https://doi.org/10.1016/j.regsus.2024.03.003.
24. Haixia Xu, Luguang Jiang, Ye Liu. Mapping the potential distribution of Asian elephants: Implications for conservation and human–elephant conflict mitigation in South and Southeast Asia. Ecological Informatics. 2024; 80(1):1-10. doi: 10.1016/j.ecoinf.2024.102518.
25. Zaw Min Thant, Peter Leimgruber, Amirtharaj Christy Williams, Zaw Min Oo. Factors influencing the habitat suitability of wild Asian elephants and their implications for human–elephant conflict in Myanmar. Global Ecology and Conservation. 2023; 43(1):1-10. doi: 10.1016/j.gecco. 2023.e02468.
26. G.B.Govindaprabhu, M. Sumathi. Ethno medicine of Indigenous Communities: Tamil Traditional Medicinal Plants Leaf detection using Deep Learning Models. Procedia Computer Science. 2024; 235(1):1135-1144. https://doi.org/10.1016/j.procs.2024.04.108.