iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: III Month of publication: March 2026

DOI: https://doi.org/10.22214/ijraset.2026.77799

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

ConvNeXt-Driven Multi-Class Animal Prediction Using Deep Feature Learning

Ketan Kanjiya[1], Piyush Sonani[2], Upendrasinh Zala[3]
[1]*Chief Research Officer, Kshatrainfotech, Ahmedabad, Gujarat, India*
[2]*Chief Technology Officer, Kshatrainfotech, Ahmedabad, Gujarat, India*
[3]*Chief Executive Officer, Kshatrainfotech, Ahmedabad, Gujarat, India*

Abstract: Automated identification of animal species from visual data is essential for biodiversity conservation, ecological monitoring, and intelligent wildlife management. This paper presents a deep learning framework for large scale multi-class animal species classification using transfer learning with the ConvNeXt-Tiny architecture. A pretrained ConvNeXt model, initialized with ImageNet weights, is fine-tuned on a publicly available dataset comprising 5,400 images across 90 distinct animal classes. The proposed system integrates modern convolutional design principles inspired by vision transformers with the efficiency and inductive biases of conventional CNNs, enabling robust hierarchical feature learning. Data augmentation and dropout regularization are employed to improve generalization and mitigate overfitting. Experimental results demonstrate strong classification performance, achieving an overall accuracy of 93%, with balanced macro precision, recall, and F1-score of 93%. Both quantitative and qualitative evaluations confirm the model's effectiveness in handling visual variability, background complexity, and inter-class similarity. The proposed framework offers a scalable and reliable solution for automated animal species recognition in real-world ecological and conservation applications.

Keywords: Multi-Class Classification, Deep Learning, Animal Classification,Transfer Learning, ConvNeXt, Computer Vision

I. INTRODUCTION

Accurate identification and classification of animal species from visual data play a crucial role in biodiversity conservation, ecological monitoring, wildlife management, and environmental sustainability. With the rapid growth of digital imaging technologies and the widespread availability of visual sensors such as cameras, drones, and remote monitoring systems, massive volumes of animal imagery are being generated daily. Manual analysis of such large scale visual data is time consuming, error prone, and infeasible at scale, creating a strong demand for automated, intelligent systems capable of recognizing animal species efficiently and reliably. Traditional computer vision approaches for animal recognition relied heavily on handcrafted feature extraction techniques, including texture descriptors, color histograms, shape features, and edge based representations. These methods, although effective in controlled environments, often struggle in real-world scenarios due to variations in lighting conditions, backgrounds, occlusions, pose changes, and intra-class diversity. Moreover, their dependence on domain specific feature engineering limits scalability and generalization across diverse species and environments. As a result, such classical approaches have gradually been replaced by data driven deep learning methodologies that learn hierarchical feature representations directly from raw image data.

In recent years, deep convolutional neural networks (CNNs) have demonstrated remarkable performance in image classification, object detection, and visual recognition tasks. Architectures such as VGG, ResNet, EfficientNet, and modern convolutional-transformer hybrids have significantly advanced the state of the art in visual learning. Transfer learning, in particular, has emerged as a powerful paradigm, enabling models pretrained on large scale datasets such as ImageNet to be adapted efficiently for domain specific tasks with limited labeled data. By leveraging pretrained feature representations, transfer learning not only accelerates convergence but also improves model generalization and robustness. Motivated by these advancements, this study proposes a deep learning based framework for automated multi-class animal species classification using a modern pretrained ConvNeXt architecture. ConvNeXt represents a new generation of convolutional networks that integrate design principles from vision transformers while preserving the computational efficiency and inductive biases of CNNs. This combination enables improved feature learning, scalability, and performance across complex visual tasks. The proposed model is fine-tuned on a large scale animal image dataset comprising 90 distinct animal classes, enabling the system to learn rich, discriminative representations for diverse species categories. This research aims to develop a reliable and scalable system for multi-class animal species identification. The proposed framework is designed to support automated species recognition under challenging environmental conditions, reducing reliance on manual annotation and accelerating data driven ecological analysis and biodiversity research.

II. LITERATURE REVIEW

The transition from traditional, manual field surveys to automated wildlife monitoring systems has been precipitated by the overwhelming volume of visual data generated by modern remote sensors [1]. While camera traps and unmanned aerial vehicles provide unprecedented access to remote habitats, the manual annotation of millions of images constitutes a significant analytical bottleneck that delays critical conservation decisions[2]. Early computer vision methodologies relied extensively on handcrafted features and shallow classifiers, such as Support Vector Machines , which frequently lacked the robustness necessary to generalize across complex, unconstrained natural environments[3,4]. However, the advent of Deep Convolutional Neural Networks has fundamentally shifted this paradigm, allowing for the automated extraction of high level discriminative features that often surpass human accuracy[5,6].

The architectural evolution of CNNs has been defined by a progressive increase in network depth and the optimization of computational efficiency[7]. Foundational models like AlexNet demonstrated the baseline efficacy of deep learning, while subsequent architectures such as VGG and GoogLeNet introduced deeper layers and optimized inception modules to refine feature representations[8]. The introduction of ResNet further addressed the vanishing gradient problem through residual learning and skip connections, enabling the training of networks with over 150 layers[9]. Recent research has increasingly turned toward hybridised models, which fuse the capabilities of networks like VGG-19 and DenseNet-121 to achieve superior precision in multi-class animal prediction tasks[10].

At the current frontier of this evolution is the ConvNeXt architecture, which represents a significant modernisation of standard convolutional networks by integrating design philosophies from modern Vision Transformers[11]. ConvNeXt maximizes the utility of multiscale feature information through parallel combinations and group convolutions, maintaining the strong inductive biases of CNNs while offering the scalability of transformer based models[11]. This architecture is particularly adept at extracting meaningful patterns from animal targets that exhibit non-rigid body structures and frequent posture changes[12]. Furthermore, its design is robust enough to mitigate the loss of fine-grained surface details that often occurs in low resolution or high noise ecological imagery[13].

A prevailing strategy in species classification is transfer learning, which involves initializing models with weights pre-trained on massive datasets such as ImageNet[2]. This methodology assumes that foundational visual features such as edges, textures, and shapes are transferable across domains, providing a significant advantage in ecological research where target datasets are often small or highly imbalanced[14]. Leveraging pre-trained knowledge allows for faster convergence and higher accuracy, even when the species in the target dataset do not overlap with the source domain[15,6]. Research indicates that transfer learning can enhance model accuracy while radically reducing the computational resources required for field deployment[9].

Despite these advancements, real-world animal identification remains confounded by significant environmental and data driven complexities[16]. Images captured in natural habitats are frequently affected by motion blur, occlusions, background clutter, and variable illumination, while nocturnal imagery often lacks the rich spectral information required for reliable visual discrimination[17, 18]. Furthermore, the dual challenges of inter class similarity, where taxonomically related species exhibit highly overlapping visual characteristics, and class imbalance, which biases learning toward dominant categories, continue to limit model generalization[19]. Recent studies suggest that two stage and hierarchical learning frameworks, which decouple detection and classification or employ specialized expert models for fine-grained recognition, represent a promising direction for scalable wildlife monitoring systems. Such architectures enable the isolation of relevant regions and the modeling of subtle morphological variations, thereby improving robustness in complex ecological environments[20]. These strategies demonstrate the growing shift toward structured, multi-level learning paradigms as a means to enhance classification reliability in large scale biodiversity analysis.

III.DATASET

Fig. 1. Representative sample images from the dataset illustrating five distinct animal categories: squirrel, lion, cow, crow, and rat.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

This study utilizes the Animal Image Dataset (90 Different Animals), a publicly available multi-class image dataset designed for large scale animal species classification tasks. The dataset consists of 5,400 labeled images distributed across 90 distinct animal categories, with each class representing a unique species. The dataset includes the following animal categories: antelope, badger, bat, bear, bee, beetle, bison, boar, butterfly, cat, caterpillar, chimpanzee, cockroach, cow, coyote, crab, crow, deer, dog, dolphin, donkey, dragonfly, duck, eagle, elephant, flamingo, fly, fox, goat, goldfish, goose, gorilla, grasshopper, hamster, hare, hedgehog, hippopotamus, hornbill, horse, hummingbird, hyena, jellyfish, kangaroo, koala, ladybugs, leopard, lion, lizard, lobster, mosquito, moth, mouse, octopus, okapi, orangutan, otter, owl, ox, oyster, panda, parrot, pelecaniformes, penguin, pig, pigeon, porcupine, possum, raccoon, rat, reindeer, rhinoceros, sandpiper, seahorse, seal, shark, sheep, snake, sparrow, squid, squirrel, starfish, swan, tiger, turkey, turtle, whale, wolf, wombat, woodpecker, and zebra. Representative sample images from different animal categories are illustrated in Figure 1.

Each image is organized into class specific directories, enabling structured supervised learning and efficient label mapping. The dataset exhibits substantial visual diversity in terms of species morphology, background complexity, and environmental context. This dataset serves as a reliable benchmark for evaluating deep learning and transfer learning models in large scale multi-class animal species classification tasks.

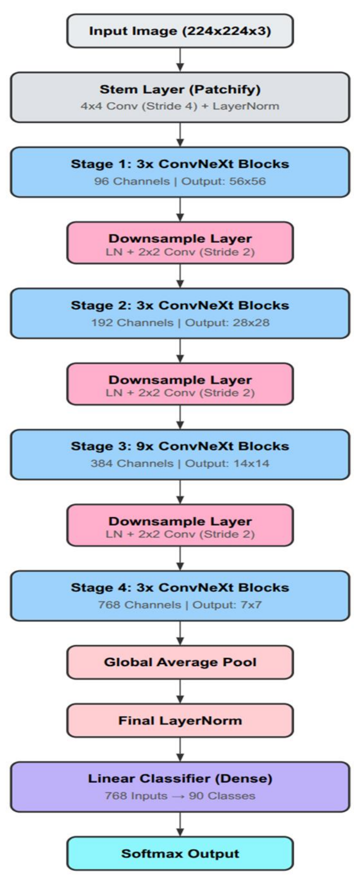
IV. METHODOLOGY

A. Data Preprocessing

All images were resized to 224 × 224 pixels, converted to RGB color space, and normalized to the range [0, 1]. Class labels were extracted from directory structures and encoded using label encoding. To enhance generalization and reduce overfitting, real-time data augmentation was applied, including random rotations, translations, shearing, zooming, and horizontal flipping. The dataset was then split into training and testing sets using an 80:20 ratio with a fixed random seed to ensure reproducibility.

B. Model Architecture

ConvNeXt-Tiny Network Architecture



International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue III Mar 2026- Available at www.ijraset.com

ConvNeXt internal block structure

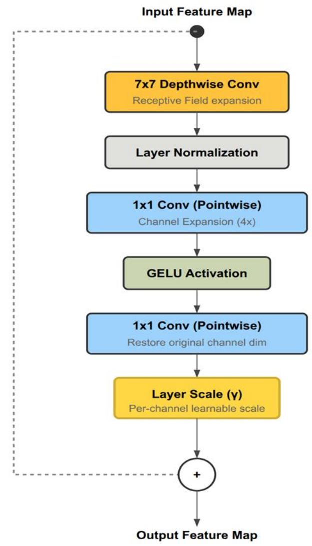


Fig. 2. ConvNeXt-Tiny network architecture and internal ConvNeXt block structure.

This study employs a transfer learning approach using the ConvNeXt-Tiny architecture pre-trained on the ImageNet dataset. ConvNeXt represents a modern convolutional neural network design that integrates architectural principles inspired by Vision Transformers while preserving the computational efficiency and inductive biases of conventional CNNs. The pre-trained ConvNeXt backbone is utilized as a deep feature extractor with the original classification head removed. Fine-tuning is performed by unfreezing the upper layers of the network to enable task specific feature adaptation. A custom classification head is then appended, consisting of a Global Average Pooling layer followed by a fully connected dense layer with ReLU activation, dropout regularization with a rate of 0.5 to mitigate overfitting, and a final softmax output layer with 90 neurons corresponding to the target animal classes. This architectural design enables hierarchical and discriminative feature learning while maintaining model stability, scalability, and generalization capability across diverse species categories.

C. Model Training

The ConvNeXt-based classification model was trained using transfer learning with ImageNet pretrained weights and fine-tuned on the target dataset. Training was performed for 20 epochs using the Adam optimizer with a learning rate of 1×10^{-4}, sparse categorical cross-entropy loss, and a learning rate scheduler for adaptive optimization. Experiments were conducted on a high performance system running Ubuntu 24.04.1 LTS, equipped with an AMD Ryzen 7 5700X CPU, 64 GB RAM, and an NVIDIA GeForce RTX 4090 GPU, enabling efficient large scale deep learning training and evaluation.

V. RESULTS AND DISCUSSION

To quantitatively evaluate the performance of the proposed model, standard multi-class classification metrics were employed, including accuracy, precision, recall, and F1-score. Accuracy measures the overall correctness of predictions, while precision evaluates the reliability of positive predictions. Recall measures the model's ability to correctly identify true instances, and the F1-score represents the harmonic mean of precision and recall, providing a balanced performance indicator.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

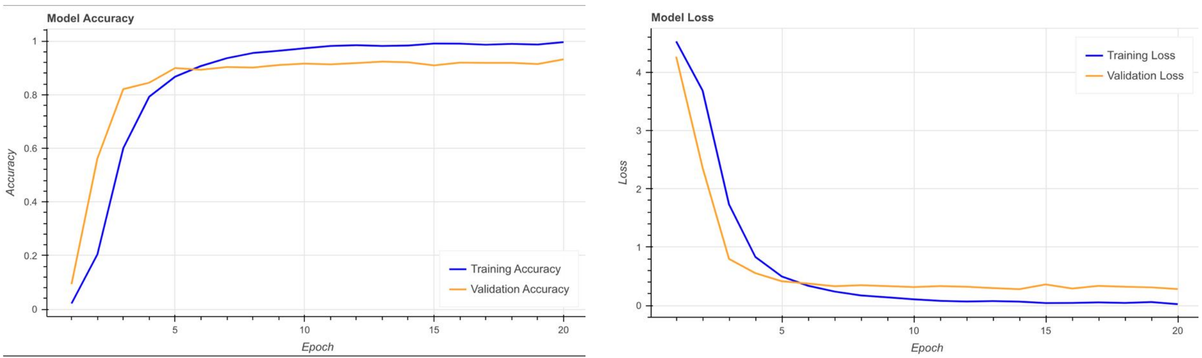
$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

A. Training Performance Analysis



(a) (b)

Fig. 3. Training dynamics of the ConvNeXt-Tiny model: (a) Accuracy vs. epochs, and (b) Loss vs. epochs, illustrating convergence behavior and generalization performance.

The training and validation accuracy and loss curves demonstrate stable convergence and effective learning behavior of the proposed ConvNeXt-based model. Accuracy consistently increases across epochs, while both training and validation losses decrease steadily, indicating successful optimization and minimal overfitting. The close alignment between training and validation trends reflects strong generalization capability and robust learning dynamics on unseen data.

B. Overall Classification Performance

Table 1: Overall Classification Performance

Metric	Value
Accuracy	93.00%
Macro Precision	93.00%
Macro Recall	93.00%
Macro F1-score	93.00%

The overall quantitative performance of the model indicates strong multi-class classification capability across the 90 animal categories. The framework achieves an overall accuracy of 93%, with macro precision, macro recall, and macro F1-score all reaching 93%, demonstrating balanced and consistent performance across classes.

C. Qualitative Prediction Analysis

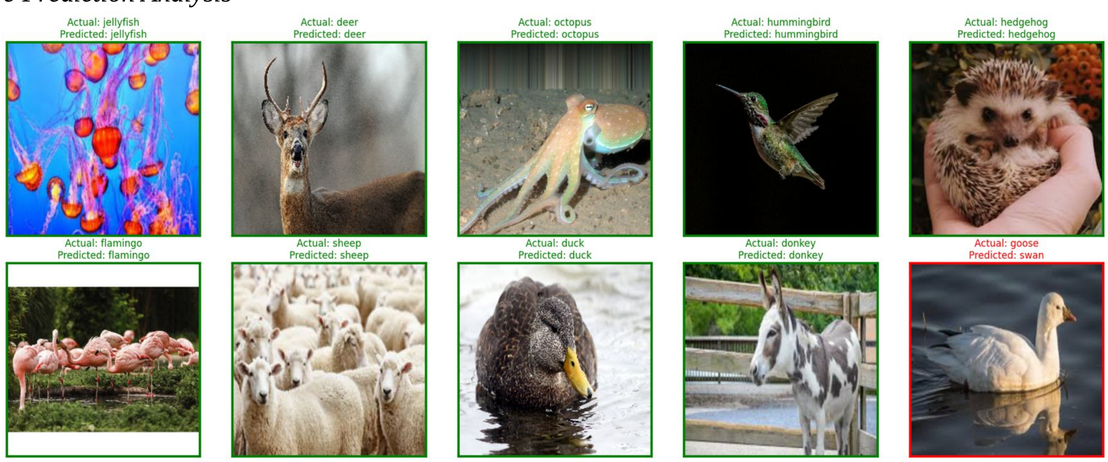


Fig. 4. Qualitative prediction results showing sample test images with corresponding ground truth (Actual) and model predicted class labels across multiple animal categories.

Qualitative evaluation through sample predictions provides visual validation of the model's discriminative capability. As illustrated in Figure 4, the predicted labels closely align with the ground truth annotations across diverse species, demonstrating strong generalization across variations in appearance, posture, and background complexity. The model consistently identifies correct species even in visually complex scenes, confirming the robustness and practical reliability of the proposed framework for real-world animal species recognition tasks.

VI. CONCLUSIONS

This study presented a deep learning framework for automated multi-class animal species classification using the ConvNeXt-Tiny architecture with transfer learning. By fine-tuning a model pre-trained on ImageNet and appending a task specific classification head, the proposed system achieved an overall accuracy of 93%, with consistently balanced macro precision, recall, and F1-score across all 90 animal categories. The results confirm that ConvNeXt's integration of transformer inspired design principles with conventional CNN inductive biases yields powerful discriminative representations, particularly in the presence of visual diversity, background clutter, and inter class similarity. Data augmentation and dropout regularization further contributed to robust generalization on unseen data. These findings demonstrate the practical viability of the proposed framework for real-world wildlife monitoring, biodiversity conservation, and ecological research applications. Future work may explore hierarchical classification strategies, larger and more imbalanced datasets, and lightweight model variants to support deployment in resource constrained field environments.

REFERENCES

[1] M. A. Rahman, T. M. Berhe, L. Borgianni, M. S. Ahmed, C. Bua, D. Adami, and S. Giordano, "A scalable framework for deploying AI-powered wildlife monitoring in resource-limited field environments," IEEE Access, vol. 13, pp. 145023–145041, 2025, doi: 10.1109/ACCESS.2025.3598927.

[2] H. Nguyen, S. J. Maclagan, T. D. Nguyen, T. Nguyen, P. Flemons, K. Andrews, E. G. Ritchie, and D. Phung, "Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring," in Proc. 2017 IEEE Int. Conf. Data Science and Advanced Analytics (DSAA), 2017, pp. 40–49, doi: 10.1109/DSAA.2017.31.

[3] W. H. S. Antônio, M. Da Silva, R. S. Miani, and J. R. Souza, "A proposal of an animal detection system using machine learning," Applied Artificial Intelligence, vol. 33, no. 13, pp. 1093–1106, 2019, doi: 10.1080/08839514.2019.1673993.

[4] A. Shalika and L. Seneviratne, "Animal classification system based on image processing and support vector machine," Journal of Computer and Communications, vol. 4, pp. 12–21, 2016, doi: 10.4236/jcc.2016.41002.

[5] W. H. E. Mg, T. T. Zin, P. Tin, M. Aikawa, K. Honkawa, and Y. Horii, "Automated cattle monitoring system for calving time prediction using trajectory data embedded time series analysis," IEEE Open Journal of the Industrial Electronics Society, vol. 6, pp. 216–234, 2025, doi: 10.1109/OJIES.2025.3533663.

[6] Z. U. Rehman, A. Syed, A. Tayab, G. G. Tejani, D. S. Khafaga, and E.-S. M. El-Kenawy, "Improving real-time animal detection using group sparsity in YOLOv8: A solution for animal-toy differentiation," Computers, Materials & Continua, vol. 86, no. 2, pp. 1–25, 2026, doi: 10.32604/cmc.2025.070310.

[7] M. B. Oliveira, H. S. Bernardino, A. B. Vieira, and D. A. Augusto, "Classification of animal species via deep neural networks and species distribution modeling: A systematic review," Artificial Intelligence Review, vol. 58, no. 8, p. 230, 2025, doi: 10.1007/s10462-024-11074-w.

[8] W. Feng, W. Ju, A. Li, W. Bao, and J. Zhang, "High-efficiency progressive transmission and automatic recognition of wildlife monitoring images with WISNs," IEEE Access, vol. 7, pp. 161412–161423, 2019, doi: 10.1109/ACCESS.2019.2951596.

[9] M. Willi, R. T. Pitman, A. W. Cardoso, C. Locke, A. Swanson, A. Boyer, M. Veldthuis, and L. Fortson, "Identifying animal species in camera trap images using deep learning and citizen science," Methods in Ecology and Evolution, 2018, doi: 10.1111/2041-210X.13099.

[10] S. S. Khan, N. V. Doohan, M. Gupta, S. Jaffari, A. Chourasia, K. Joshi, and B. Panchal, "Hybrid deep learning approach for enhanced animal breed classification and prediction," Traitement du Signal, vol. 40, no. 5, pp. 2087–2099, 2023, doi: 10.18280/ts.400526.

[11] A. He, X. Li, X. Wu, C. Su, J. Chen, S. Xu, and X. Guo, "ALSS-YOLO: An adaptive lightweight channel split and shuffling network for TIR wildlife detection in UAV imagery," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 17308–17326, 2024, doi: 10.1109/JSTARS.2024.3461172.

[12] R. A. Rajagukguk, S.-Y. Lee, J.-Y. Park, K. F. Daniel, C.-R. Lee, Z. Chen, D. Liu, T. Norton, J. Park, and S.-W. Hong, "Deep learning for visual animal monitoring (detection, tracking, pose estimation, and behavior classification): A comprehensive review," Smart Agricultural Technology, vol. 12, p. 101539, 2025, doi: 10.1016/j.atech.2025.101539.

[13] Y. Wu, X. Liu, J. Hao, Z. Xing, H. Xi, W. Sun, and Q. Yao, "MBAF-MSCNet: Multi-branch fusion for multi-scale context in infrared wildlife detection," IEEE Access, vol. 13, pp. 165830–165843, 2025, doi: 10.1109/ACCESS.2025.3605738.

[14] H. S. de A. Martins, G. S. Ferrante, and R. I. Meneguette, "Descriptor: A Brazilian roadkill animals for object detection GAN-augmented dataset (BRAGAN)," IEEE Data Descriptions, pp. 1–10, 2026, doi: 10.1109/IEEEDATA.2026.3664029.

[15] A. Negreiro, R. E. P. Ferreira, T. Bresolin, G. L. Menezes, E. Casella, D. Döpfer, L. L. Hernandez, G. J. M. Rosa, and J. R. R. Dórea, "Artificial intelligence for animal identification: Development of a computer vision system for reliable dairy cattle traceability across developmental phases," Journal of Dairy Science, 2026, doi: 10.3168/jds.2025-27653.

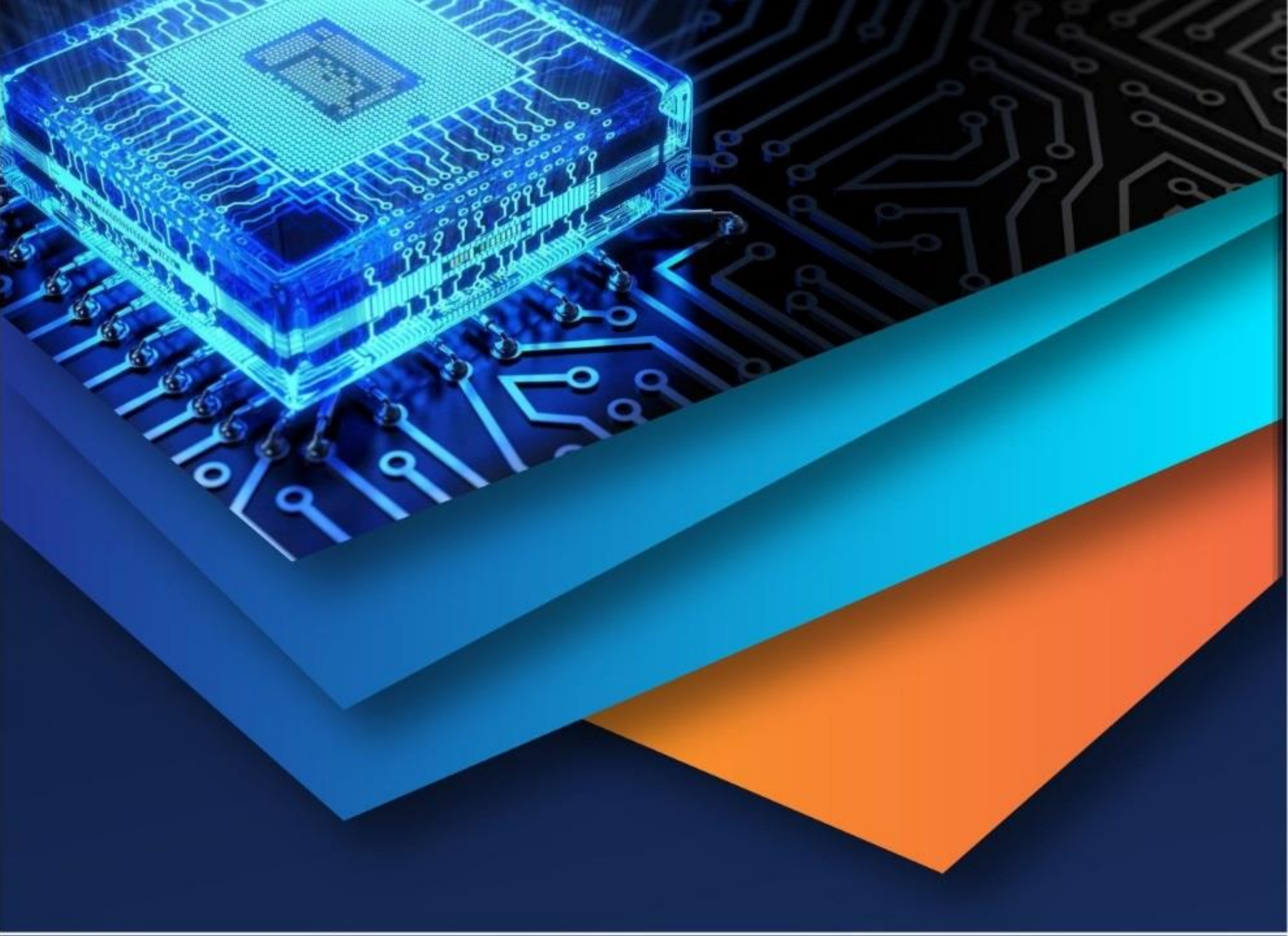
[16] Q. Zhang, K. Ahmed, N. Sharda, and H. Wang, "A comprehensive survey of animal identification: Exploring data sources, AI advances, classification obstacles and the role of taxonomy," International Journal of Intelligent Systems, vol. 2024, no. 1, p. 7033535, 2024, doi: 10.1155/2024/7033535.

[17] A. B. Mughal, R. U. Khan, A. U. Rehman, and A. Bermak, "Deep learning for dynamic wildlife monitoring: A real-time approach," IEEE Access, vol. 13, pp. 147422–147448, 2025, doi: 10.1109/ACCESS.2025.3600625.

[18] K. Parkavi, A. Ganguly, S. Sharma, A. Banerjee, and K. Kejriwal, “Enhancing road safety: Detection of animals on highways during night,” IEEE Access, vol. 13, pp. 36877–36896, 2025, doi: 10.1109/ACCESS.2025.3545490.

[19] T. Battu and D. S. R. Lakshmi, “Animal image identification and classification using deep neural networks techniques,” Measurement: Sensors, vol. 25, p. 100611, 2023, doi: 10.1016/j.measen.2022.100611.

[20] M. Mulero-Pázmány, S. Hurtado, C. Barba-González, M. L. Antequera-Gómez, F. Díaz-Ruiz, R. Real, I. Navas-Delgado, and J. F. Aldana-Montes, “Addressing significant challenges for animal detection in camera trap images: A novel deep learning-based approach,” Scientific Reports, vol. 15, no. 1, p. 16191, 2025, doi: 10.1038/s41598-025-90249-z.



10.22214/IJRASET

INDEX COPERNICUS
45.98

IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR:
7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY