A Comprehensive Survey of AI enabled Techniques for Automated Animal Detection and Classification in Ecological and Real-Time Applications

Muskan[1], Sinchana G P[2], Prarthana B G[3], Karthik M L[4], Arjun B C[5]
[1,2,3,4,5]*Malnad College of Engineering*

***Abstract*—Artificial intelligence (AI) is coming to play a fundamental role in responding to ecological issues, especially wildlife conservation. This review consolidates the use of innovative AI methods for the automated detection, classification, and tracking of animals in various environments, i.e., forests, farmland, highways, and poorly lit situations. The work points out key gains in accuracy and real-time processing through these technologies. Principal innovations are techniques for transcending data constraints, combination of AI with IoT devices for distant monitoring, and sophisticated segmentation techniques for improved detection accuracy. The applications of AI go beyond technical innovation to real-world implications like the mitigation of human-wildlife conflict, sustainable agriculture, and ecosystem management. Through combining existing progress and recognizing nascent trends and challenges, this paper presents significant lessons and insights for future studies aimed at utilizing AI in the promotion of ecological sustainability and efficient animal monitoring.**

***Index Terms* —Animal Detection, Deep Learning, Wildlife Monitoring, Human-Wildlife Conflict, Smart Agriculture**

I. INTRODUCTION

With the rapid advancement of artificial intelligence (AI) and deep learning, wildlife monitoring and animal detection have undergone a technological transformation. The growing concerns over biodiversity loss, habitat encroachment, and human-wildlife conflict demand innovative and automated solutions for efficient ecological surveillance. AI-driven systems, particularly those leveraging convolutional neural networks (CNNs) and object detection models like YOLO, Mask R-CNN, and ResNet, have demonstrated exceptional capability in recognizing and classifying animals in complex environments such as forests, agricultural zones, highways, and wildlife sanctuaries.

This literature survey explores recent breakthroughs in the field, examining a wide array of methods that combine deep learning, computer vision, and IoT to detect, classify, and track animals across diverse ecological and real-time scenarios. The reviewed works cover innovations in pseudo-labeling, genetic segmentation, sensor integration; each contributing to the creation of smart, scalable, and robust wildlife monitoring frameworks. Through this comprehensive review, we aim to provide insights into the state-of-the-art models and techniques that are shaping the future of animal detection, with a special focus on their accuracy, adaptability, and practical applicability in conservation and safety domains.

II. LITERATURE REVIEW

The review of literature on AI-based methods for animal detection and classification is essential because of the emerging demand for precise, real-time solutions in a wide range of applications including agriculture, wildlife conservation, road safety, smart surveillance, and prevention of human-wildlife conflict. Conventional methods are time-consuming and not very efficient in changing environments like forests, farms, and highways. AI algorithms, particularly CNNs, YOLO, and Mask R-CNN, alongside IoT and edge computing, provide scalable, automated systems for applications such as crop protection, species monitoring, and intrusion detection. This review analyzes existing solutions, emphasizes model performance, and determines research gaps, providing insights to inform the design

of effective, adaptive solutions that promote ecological sustainability and public safety.

2.1 Agriculture and Livestock Management

The literature review on Agriculture and Livestock Management is crucial due to increasing challenges like crop damage, livestock threats, and inefficient traditional monitoring. AI-driven systems, especially those using CNNs and IoT, offer real-time, automated solutions for detecting and responding to animal activity. Reviewing existing work helps identify effective models, such as genetic segmentation and bioacoustic monitoring, and highlights performance trends and gaps. This understanding supports the development of smarter, more sustainable approaches to protect agriculture and manage livestock efficiently.

Kuei-Chung Chang et al. [1] developed a IoT-powered agricultural warning system to counter Taiwan macaque-induced crop damage, which evolves into conventional deterrents such as noise-makers or traps. The system utilizes IP cameras and a hybrid image recognition algorithm consisting of a preprocessing (background subtraction and colour masking), a feature mapping (FAST algorithm and k-means clustering), and a verification (LBP cascade classifier) module to identify monkeys in real time. Upon sensing, the system initiates alarms (e.g., sound) and notifies farmers through MQTT-based IoT architecture and sends data to a server for processing. Validation with 134 monkey samples realized 91% recognition rate (122 true positives, 12 false negatives) and 15% false positives. Created with scalability and integration into larger smart agriculture systems in mind, the solution overcomes the inefficiency of traditional approaches and provides farmers with a proactive, automated method of crop protection.

L. G. C. Vithakshana et al. [2] presented “IoT based animal classification system using convolutional neural network”. The paper presents an IoT-based system for classifying animal sounds using convolutional neural networks (CNNs) to aid ecological monitoring. The authors collected 417 two-second audio clips of ten animal species (e.g., bat, elephant, hornbill) from online libraries, preprocessed them using Mel-frequency Cepstral Coefficients (MFCC) for feature extraction, and trained a CNN model with two convolutional layers, ReLU activation, and max-pooling. Testing six optimizers, AdaDelta, Gradient Descent, and RMSProp achieved the highest accuracy of 91.3%, outperforming prior studies with imbalanced datasets. A hardware module (Arduino UNO, NodeMCU, and microphone) captured sounds in remote locations, transmitting data to a cloud server for analysis via an Android app. The system successfully classified species in real-world tests, demonstrating its utility for bioacoustics monitoring. Future work includes expanding the dataset, refining hyperparameters, and extending classification to animal gender, age, and mood detection. The study highlights CNNs' effectiveness for acoustic animal classification and IoT's role in remote ecological research.

Yong Jae Lee et al. [3] presented “Using pseudo-labeling to improve performance of deep neural networks for animal identification”. This paper investigates the use of pseudo-labeling, a semi-supervised learning (SSL) technique, to enhance the performance of deep neural networks (DNNs) for identifying individual Holstein cows in camera-trap images. By leveraging a small manually labeled dataset (2,354 images) and a larger unlabeled dataset (20,194 images), the study demonstrates that pseudo-labeling iteratively training models on confident predictions from unlabeled data can improve identification accuracy by up to 20.4 percentage points, achieving 92.7% accuracy on a test set of 59 cows. The optimal confidence threshold for pseudo-labeling was found to be 0.999, balancing label quality and dataset size. The Xception architecture outperformed others (MobileNetV2, NASNet Large) in efficiency and accuracy. The method significantly reduces manual labeling effort, showing promise for scaling to large herds and other agricultural applications. The study highlights SSL’s potential to address data scarcity in animal identification while maintaining high performance. Data and code are available in an Open Science repository

Yang Zhao et al. [4] proposed “Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: AReview”. The paper provides a comprehensive review of how convolutional neural networks (CNNs) are applied in precision livestock farming. It categorizes and analyzes 105 publications, focusing on five core computer vision tasks image classification, object detection, semantic/instance segmentation, pose estimation, and tracking applied across major livestock species such as cattle, pigs, sheep/goats, and poultry. The study highlights critical aspects of system

development, including camera setup, image type, data labeling, GPU selection, and preprocessing techniques. It emphasizes the benefits of CNNs for non-invasive, real-time monitoring of animal health, behavior, and welfare, and outlines challenges such as data imbalance, occlusion, lighting variability, and the need for high-quality labeled data. Performance analysis shows that over 60% of reviewed models achieved accuracy above 90%, with some reaching up to 100%, though issues like overfitting were noted. The review concludes with future research directions, calling for more robust algorithms, better datasets, interdisciplinary collaboration, and the integration of CNNs with other AI technologies to enhance productivity and animal welfare in smart farming systems.

Davide Adami et al. [5] proposed a system that overcomes the problem of crop destruction by ungulates (e.g., wild boars, deer) in rural regions through the creation of an energy-efficient, AI-based system that incorporates Edge-AI, IoT, and computer vision to identify and deter animals non-lethally. The system utilizes PIR sensors for detecting motion and cameras for real-time animal detection using YOLOv3 and Tiny-YOLOv3 models, complemented with species-specific ultrasound emissions (18–27 kHz) that are inaudible to humans. Tested on low-power edge devices (NVIDIA Jetson Nano, Raspberry Pi 3B+ with/without Intel Movidius NCS), the work identified NVIDIA Jetson Nano executing Tiny-YOLOv3 as optimal, with a rate of 15 FPS and accuracy of 82.5% mAP, trading power efficiency (1.2 FPS/W) and cost (6.6 USD/FPS) for performance, beating out Raspberry Pi configurations that were subjected to thermal throttling (93°C peaks). Taking advantage of LoRaWAN for rural connectivity and cloud integration, the solution showcases a scalable, environmentally friendly method of minimizing crop loss, with future development focusing on reliability enhancements and wider agricultural uses through embedded Edge-AI innovation.

G. Nagarajan et al. [6] presented “DeepAID: a design of smart animal intrusion detection and classification using deep hybrid neural networks”. This paper presents DeepAID, a smart animal intrusion detection system that uses hybrid deep convolutional neural networks (CNNs) combined with IoT and image processing to identify and deter wild animals from entering restricted areas like farms. The system employs camera traps to capture real-time images, processes them using a multi-layer CNN architecture (convolution, pooling, and fully connected layers), and triggers species-specific alarms (e.g., predator sounds) to repel intruders. Tested on datasets including Amur Tiger Re-identification (ATRW), Animals Detection Images Dataset (ADID), and Google Open Images V6+, the model achieved 99.6% accuracy on ADID and 95.6% on Google Open Images, with high precision (0.93–0.99) and F1-scores (0.95–0.99). The system, optimized for Raspberry Pi deployment, effectively addresses human-wildlife conflicts by minimizing crop damage and accidents while operating robustly in day/night conditions. Future enhancements aim to expand species coverage, integrate cloud-based scalability, and refine repellent mechanisms for broader conservation applications.

Ramakant Chandrakar et al. [7] designed a system for animal detection based on deep learning with a genetic algorithm for segmentation and a 25-layer Convolutional Neural Network (CNN) for classification. The authors solve problems in wildlife monitoring, including high false positives/negatives in current approaches, by incorporating genetic algorithms to optimize image segmentation thresholds, improving feature extraction from saliency maps. Trained on the ECSSD dataset of 1,000 images of animals and non-animals, their model demonstrates state-of-the-art performance with 99.02% precision, 98.79% recall, 98.9% F-Measure, and 0.78% MAE compared to traditional methods such as SU, DS, and MDF. CNN architecture consists of convolutional, ReLU, pooling, and fully connected layers and processes resized images (227×227×3) divided into 80% training and 20% test. Applications in practice include collision avoidance and wildlife protection via real-time detection with webcams. The research demonstrates the strength of the integration of genetic algorithms with deep learning to ensure strong animal detection in real-world environments, and future studies would focus on the inclusion of other features to continue improving accuracy.

Apirak Sang-ngenchai et al. [8] developed a deep learning-based system for preventing Japanese macaque monkeys from destroying sweet potato crops in Hakusan, Japan, where losses to agriculture through wildlife are large. The system uses the YOLOv4 Tiny algorithm for real-time object detection trained on 2,580 images annotated from trail cameras and

implemented with RTSP-enabled cameras to observe fields. Using a high-performance Windows PC with Nvidia RTX 3080 GPU, the model scored precision, recall, and AP@0.5 of 0.7310, 0.8462, and 0.7421, respectively, during k-fold cross-validation. Solar power-based hardware with weather-resistant huts and a recharging lithium-ion battery allows 24/7 usage, while real-time notification through the Line application allows farmers to immediately react to intrusions. Although there are sometimes misclassifications from image noise or range, the system is efficient in mitigating crop damage. Future development will focus on broadening the dataset to other wildlife and incorporating IoT technologies for wider agricultural monitoring use.

C. Thilagavathi [9] proposed an AI-driven intelligent animal tracking and detection system with YOLOv8 for real-time object recognition and IoT devices for multi-dimensional alerts. The system uses YOLOv8, which is trained on a wildlife images dataset (divided into 70% training, 15% validation, and 15% test sets), for animal detection in various environments. The YOLOv8x model performed the best with a mean Average Precision (mAP) of 94.3%, precision of 91.0%, and recall of 89.9%, showing a detection efficacy of 96%. The system invokes context-dependent reactions, including buzzers to discourage intruders and automated SMS/email notifications to authorities, upon the identification of threats such as poaching. IoT integration facilitates real-time tracking and quick intervention, with the ability to adapt to environments such as forests and highways. The paper points out an accuracy-recall trade-off but stresses the strong robustness of the model for balancing speed and accuracy, well-suited to wildlife conservation as well as to conflict mitigation.

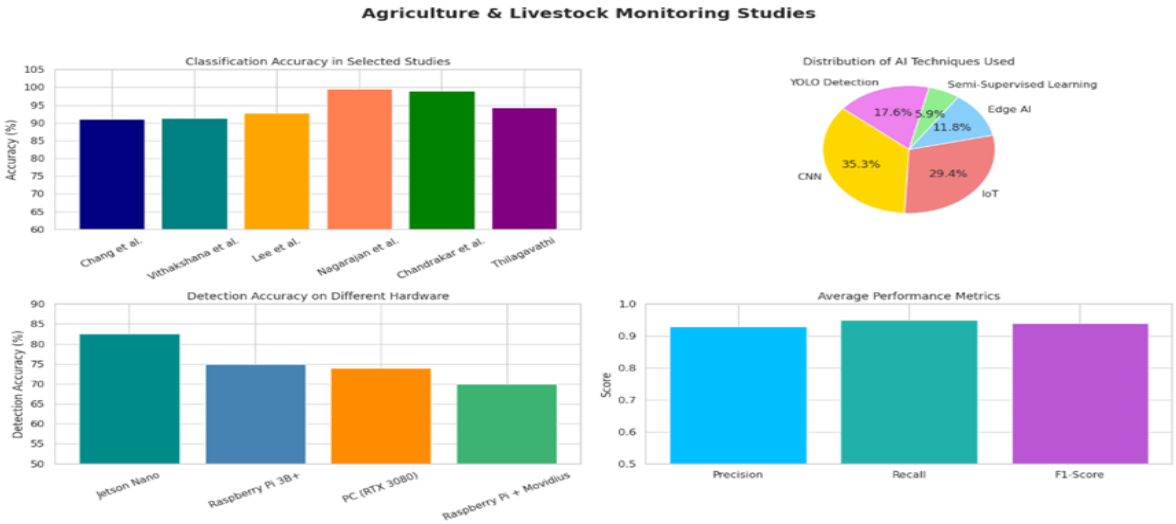

Fig 2.1: Overview of AI Techniques and Performance Metrics in Agriculture and Livestock Monitoring

Fig 1 offers a glimpse into visualization of AI use for agricultural animal monitoring. The top-left bar graph of classification accuracies between different studies highlights Nagarajan et al. and Chandrakar et al. having remarkable results above 98%. The use of CNN and IoT methods is displayed as a pie chart, of which CNN is the most prevalent. The bottom-left bar graph spotlights the contribution of hardware to detection accuracy, with Jetson Nano being the best performer. The bottom-right bar chart aggregates mean performance values (precision, recall, F1-score), all > 0.9, indicating robust model reliability. This set of charts highlights how AI can heavily contribute toward efficient livestock and agricultural monitoring.

2.2 Wildlife Conservation and Monitoring

The literature review on Wildlife Conservation and Monitoring is important as it addresses the need for accurate, scalable, and non-invasive methods to monitor animal populations and habitats. Traditional techniques are often limited by manual labor and environmental constraints. AI-powered models, particularly CNNs and object detection frameworks like YOLO and ResNet, have shown high accuracy in identifying species from camera trap images, even in challenging conditions. Reviewing these

advancements helps evaluate their effectiveness, identify technological gaps, and guide future innovations, ultimately supporting biodiversity conservation and ecological research.

Alexander Gomez et al. [10] proposed "Towards Automatic Wild Animal Monitoring: Identi cation of Animal Species in Camera-trap Images using Very Deep Convolutional Neural Networks". This paper proposes an automatic approach to identifying animal species in camera-trap images using deep convolutional neural networks (ConvNets). The study addresses the challenges posed by unbalanced datasets, environmental conditions, and varying animal poses by experimenting with four dataset versions from the Snapshot Serengeti dataset and employing six ConvNet architectures. The proposed method, using models like ResNet-101, achieves high accuracy (88.9% Top-1 and 98.1% Top-5) when trained on balanced and manually segmented data, outperforming previous approaches. The paper discusses the limitations related to image quality, dataset imbalance, and fine-grained classification and highlights the need for improved segmentation algorithms and more diverse data to enhance automation in wildlife monitoring.

Hayder Yousif et al. [11] came up with a way to efficiently and effectively identify animals and people within complicated camera-trap photos captured in the wild. Their system involves a two-stage process: it first removes the background from the images through a process known as low-rank and sparse decomposition, which eliminates areas of the image that don't move (such as trees or rocks). Next, a deep learning model (a convolutional neural network) scans the rest of the image to determine if it holds an animal, a human, or nothing at all. This approach serves to decrease the quantity of unnecessary information the model needs to process, resulting in the detection being quicker and more accurate. The findings indicated that this technique is effective even in obstructed spaces and could be applied beneficially in real-time wildlife monitoring and conservation practice.

Adrian Gomez Villa et al. [12] introduced a deep learning approach to automatically identify animal species using camera-trap data. They trained convolutional neural networks such as VGGNet and AlexNet on a dataset containing images of 26 animal species under various lighting and occlusion conditions. The system included preprocessing steps like image resizing and normalization to improve learning stability. The final model achieved a classification accuracy of 92%, even when animals were partially hidden or in motion. This research supports the integration of CNNs into wildlife ecology and conservation technologies for fast, accurate species identification.

Macro Willi et al. [13] designed a system which is integration of citizen science and deep learning to efficiently classify animal species in camera trap photos. With convolutional neural networks (CNNs) trained on datasets labelled by volunteers through the Zooniverse platform, the researchers tested four datasets from different regions (Tanzania, South Africa, Gabon, and the U.S.), with varying size and species composition. They used ResNet18 architecture, training individual models to differentiate empty images and species, and applied transfer learning from the largest dataset (Snapshot Serengeti) to improve performance on smaller datasets. Results were high accuracies: 88.7–92.7% for species identification and 91.2–98.0% for empty images, with transfer learning enhancing accuracy by up to 10.3% for smaller datasets. Confidence thresholding also improved accuracy to near-human levels by removing low-confidence predictions. A real-time experiment that merged model predictions with citizen scientist annotations eliminated 43% of human effort while preserving 99.78% accuracy. Yet, the rare species were challenging with limited training data. The research shows that CNNs, supplemented by citizen science, substantially reduce processing time of large camera trap datasets, making scalable ecological monitoring feasible despite difficulties in dealing with rare species.

Jason Parham et al. [14] developed a five-stage deep learning pipeline for the automatic detection and identification of animals intended to improve wildlife censusing and conservation. The pipeline consists of: 1) image classification for detecting species presence, 2) YOLO-based annotation localization (81.67% mAP), 3) species and viewpoint classification (94.28% and 87.11% accuracy, respectively), 4) background segmentation for separating animal areas, and 5) a new "Annotation of Interest" (AoI) classifier (72.75% accuracy) favoring recognisable subjects. Evaluated on the Wildlife Image and Localization Dataset (WILD) which consists of 5,784 real-world images with 12,007 annotations from 28 species the system

performs well in difficult situations such as overlapping herds and visually similar sub-species (e.g., giraffes and zebras). The major findings are enhanced localization performance (90.62% mAP) when concentrating on AoIs and high image classification accuracy (mean AUC 98.27%). The pipeline lightens computational burden by removing non-critical annotations, enabling scalable ecological surveillance. Future research plans to improve AoI classification and increase the dataset.

Mohammad Sadegh et al. [15] presented "Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning". This paper demonstrates the application of deep learning to automate the analysis of wildlife camera-trap images from the Snapshot Serengeti dataset, achieving high accuracy in species identification (93.8% top-1, 99.1% top-5), animal counting (63.1% exact count, 84.7% within ±1 bin), and behavior description (76.2% accuracy). By employing a two-stage pipeline first detecting animals (96.8% accuracy) and then extracting detailed information the system matches human volunteer performance (96.6% accuracy) while automating 99.3% of the labeling process, saving over 17,000 hours of manual effort for the 3.2 million-image dataset. The study highlights the potential of transfer learning for smaller projects, enabling significant automation even with limited labeled data. Challenges include handling rare species and multi-species images, which remain areas for future improvement. Overall, this approach transforms camera-trap data into actionable ecological insights, reducing costs and accelerating wildlife research and conservation efforts.

Norouzzadeh et al. [16] developed an automated animal identification system using deep convolutional neural networks, such as ResNet-50 and Inception-ResNet-v2, to process over 3 million camera trap images. Their method focused on two tasks: detecting the presence of an animal and classifying its species. They utilized transfer learning to overcome the issue of class imbalance and improve training with limited labeled data. The model achieved 96.6% accuracy in detecting animal presence and 93.8% in classifying species, which significantly reduced manual labor in ecological studies. Their work demonstrated how deep learning can revolutionize wildlife monitoring on a global scale with scalable and highly accurate systems.

Kylie Andrews et al. [17] described "Animal Recognition and Identification with Deep Convolutional Neural Networks for Automated Wildlife Monitoring". This paper presents a deep learning framework for automating wildlife monitoring using camera trap images, focusing on detecting and identifying animals in the wild. The authors leverage a dataset from the Wildlife Spotter project, which includes images labeled by citizen scientists, to train convolutional neural networks (CNNs) for two tasks: (1) detecting whether an image contains an animal (binary classification) and (2) identifying the species (multiclass classification). The study evaluates three CNN architectures Lite AlexNet, VGG-16, and ResNet-50 on both balanced and imbalanced datasets. Results show high accuracy, with 96.6% for animal detection (using VGG-16) and 90.4% for identifying the three most common species (using ResNet-50). The framework demonstrates robustness to data imbalance and highlights the potential of deep learning to streamline wildlife monitoring by reducing reliance on manual annotation. Future work includes improving performance through data enhancement and integrating the system as a recommendation tool for citizen science projects.

Kellenberger et al. [18] came up with a system using the Single Shot Multibox Detector (SSD) for detecting animals like elephants, giraffes, and zebras in aerial drone footage. Their pipeline processed high-resolution images captured from UAVs flying over African savannahs and included pre-trained SSD networks fine-tuned for animal shapes and movement patterns. The model achieved a detection accuracy of 83% while maintaining real-time inference at 30 FPS. This method proved effective even in scenarios with overlapping animal groups or challenging terrain and was proposed as a reliable solution for monitoring large areas with minimal human involvement.

Thirupathi Battu et al. [19] presented "Animal image identification and classification using deep neural networks techniques". This paper describes a deep learning approach for identifying and classifying animal species from camera-trap images, focusing on handling noisy labels in datasets. The authors propose a robust learning strategy that divides training data into groups using k-means clustering and trains multiple networks on these groups. The method employs maximum voting to correct noisy labels and improve

classification accuracy. The study evaluates the approach on two public datasets, Snapshot Serengeti and Panama-Netherlands, with varying noise levels (30%, 50%, and 70%). Results demonstrate that the method achieves higher accuracy compared to traditional CNNs, with 73.09% accuracy at 30% noise and 59.66% at 50% noise. The framework is particularly useful for citizen science projects where noisy labels are common. The study highlights the importance of network diversity and iterative label refinement for improving classification performance in wildlife monitoring applications. Future work could explore integrating this method with real-time systems for broader ecological research.

John K. Khaemba et al. [20] designed a YOLOv3-based detection model to identify and count large herbivores in drone-captured images from African conservation sites. Their workflow involved enhancing aerial images for contrast and feeding them into a YOLOv3 architecture trained on thousands of annotated wildlife images. The model achieved a precision of 89% and recall of 85%, delivering accurate results even at high altitudes and varying scales. The system operated at real-time speeds (~45 FPS), making it well-suited for automated wildlife censuses and continuous monitoring over wide terrains.

Dr.R.S. Sabeenian [21] proposed a convolutional neural network (CNN)-based wild animal intrusion detection system for avoiding human-wildlife conflict in proximity to forests. The method uses CNN to detect images or video frames obtained using surveillance systems as animal or non-animal objects. Transfer learning using pre-trained models like MobileNet or ResNet is utilized for enhanced detection accuracy with less training data. As soon as an intruder is identified, an alert is triggered and forwarded to officials or local groups through SMS or other IoT-supported notification systems. The model indicated high accuracy for animal classification and was effective in real-time monitoring applications, opening up the possibility of its application in forest boundary areas to further safety and monitoring.

Frank Schindler et al. [22] developed “Identification of animals and recognition of their actions in wildlife videos using deep learning techniques”. This paper presents a deep learning-based pipeline for detecting and classifying animals and their actions in wildlife videos captured by infrared camera traps. The study focuses on four animal classes deer, boars, foxes, and hares recorded mostly at night near a highway bridge in Bavaria, Germany. The authors evaluate two object detection methods, Mask R-CNN and Flow-Guided Feature Aggregation (FGFA), with Mask R-CNN achieving superior results (63.8% average precision). For action recognition, they compare three ResNet variants and the SlowFast architecture, finding that the (2+1) D ResNet performs best (94.1% accuracy). The fused pipeline combines Mask R-CNN for detection and (2+1) D ResNet for action recognition, demonstrating a proof-of-concept for automated wildlife monitoring. The study highlights the potential of this approach for biodiversity conservation while noting challenges like articulated animal movements and suggesting future improvements such as action detection and individual re-identification.

Michael R. J. Forstner et al. [23] presented “Animal Species Recognition with Deep Convolutional Neural Networks from Ecological Camera Trap Images”. The paper explores the use of deep learning to classify herpetofaunal species (toads/frogs, lizards, and snakes) from camera trap images, addressing challenges like imbalanced datasets and small body sizes. The authors compared a self-trained CNN (CNN-1) with pretrained models VGG16 and ResNet50, finding that VGG16 achieved the highest multiclass accuracy (87%), followed by ResNet50 (86%) and CNN-1 (72%). Binary classification outperformed multiclass, with toads identified most accurately (96-99%) and lizards posing the greatest challenges due to their size and camouflage. Data augmentation improved pretrained models but hindered CNN-1, likely due to limited data. The results highlight the superiority of transfer learning for ecological monitoring, though human oversight remains crucial. Future work should expand datasets and explore advanced architectures like YOLO or EfficientNet to enhance performance.

Charles M Francis et al. [24] came up with the “Using Web images to train a deep neural network to detect sparsely distributed wildlife in large volumes of remotely sensed imagery: A case study of polar bears on sea ice”. This paper explores the feasibility of using web-sourced images to train a deep neural network (ResNet-50) for detecting sparsely distributed wildlife, specifically polar bears, in large volumes of remotely sensed aerial imagery. Due to the scarcity of polar bears in the target aerial survey dataset (only 21

confirmed images out of ~61,000 photos), the researchers harvested and edited 534 web images of polar bears to resemble the aerial survey images, combining them with 6,292 background images for training. The trained CNN achieved a 95% detection rate for polar bears and a 99.4% accuracy in classifying background images, demonstrating that web-sourced images can effectively train CNNs for wildlife detection in challenging, data-scarce scenarios. The study also discusses workflows to manage false positives and highlights broader applications for monitoring rare species in remote environments.

Gyanendra K and Pragya Gupta [25] proposed "Wild Animal Detection Using Deep Convolutional Neural Network". This paper presents a robust method for wild animal detection in highly cluttered camera-trap images using deep convolutional neural networks (DCNNs). The authors address challenges such as dynamic backgrounds, varying illumination, and animal camouflage by employing a verification model that combines spatiotemporal region proposals with DCNN features. The proposed system uses Iterative Embedded Graph Cut (IEGC) to generate candidate animal regions and fine-tuned DCNN features for classification, achieving an accuracy of 91.4% with weighted KNN. The model is evaluated on a standard camera-trap dataset containing 20 animal species, demonstrating superior performance over existing methods in terms of recall (98.25%), precision (91.6%), and F1-score (0.9476). The study highlights the effectiveness of self-learned DCNN features and machine learning algorithms like SVM and ensemble classifiers for wildlife monitoring, offering a scalable solution for automated animal detection in dynamic natural environments. The system's robustness to pose variations and cluttered scenes makes it suitable for both daytime and nighttime monitoring applications.

Matthew T. Duggan et al. [26] proposed a transfer learning-based CNN model to automate camera trap image processing with minimal human effort, overcoming labor-intensity issues of manual analysis in ecological research. With the Faster R-CNN Inception v2 model, the method was trained on a small dataset (average 275 labeled images per species) from a 170-camera network in South Carolina, for 17 species classes (e.g., deer, coyote, armadillo). Bounding box labelling using Labelling software supported supervised training, and transfer learning minimized the amount of data needed. Tested on 5,277 images (90:10 train-test split), the model performed at 92% accuracy and 85% average F1 score, with validation on 10,983 images registering 93% accuracy and 76% F1 score. Pivotal innovations involved favouring F1 score over accuracy to counterbalance true-negative bias and minimizing confidence thresholds (0.90 CT). The approach evidenced species-specific efficacy (e.g., 98% accuracy for armadillos) and remained feasible for small studies, lessening dependence on large datasets. Applied to an Nvidia 2070 Super GPU, the solution provides cost-efficient, scalable wildlife monitoring with ecological relevance for biodiversity conservation.

Li Zhang et al. [27] came up with a real-time animal detection system based on YOLOv5, targeting camera-trap images taken in dense forest environments. Their method involved fine-tuning YOLOv5 on a large wildlife dataset with images of deer, foxes, boars, and other animals. To enhance model generalization, various data augmentation techniques like rotation, scaling, and brightness adjustment were used. The trained model achieved a mean Average Precision (mAP) of 87.4% and operated at 50 frames per second (FPS), proving highly efficient in detecting animals in real-time, even in low-light and cluttered scenes. This approach significantly reduced false positives and demonstrated strong potential for practical applications in forest surveillance and biodiversity tracking.

Kristina Ran ci et al. [28] proposed "Animal Detection and Counting from UAV Images Using Convolutional Neural Networks". This paper presents a method for detecting and counting deer in UAV (drone) images using convolutional neural networks (CNNs), specifically comparing the performance of YOLOv3, YOLOv4, YOLOv4-tiny, and SSD models. The study focuses on automating wildlife monitoring in the Plavna hunting ground in Serbia, where manual counting is time-consuming and error-prone. The dataset consists of 30 high-resolution UAV images, which were divided into smaller tiles (416 × 416 pixels) to improve detection accuracy, resulting in 2340 images with 169 manually annotated deer. The models were evaluated based on mean average precision (mAP), precision, recall, and F1 score, with YOLOv4 achieving the highest mAP of 70.45% and a precision of 86%, while YOLOv4-tiny offered faster real-time performance with slightly lower accuracy.

The counting function applied to the best-performing models yielded errors of 8.3% (YOLOv4) and 7.1% (YOLOv4-tiny). The study highlights the challenges of detecting small objects in cluttered environments and addresses issues like class imbalance and dataset diversity. The results demonstrate the potential of UAVs and deep learning for efficient and accurate wildlife monitoring, with future work suggesting improvements such as higher-resolution models and expanded datasets for broader applications.

Mengyu Tan et al. [29] proposed three popular deep learning models (YOLOv5, FCOS, Cascade R-CNN) for wildlife detection and classification from China's Northeast Tiger and Leopard National Park (NTLNP) camera trap images. The authors built the NTLNP dataset, which contains 25,657 annotated images of 17 species (rare Amur tigers and leopards among them), divided into day (15,313) and night (10,344) subcategories. Joint training of models on day-night data performed better than individual training, with the best accuracy attained by YOLOv5m : 98.9%. The models performed well for large, clear species but did poorly for small, blurry, or occluded animals. On video classification, YOLOv5m retained strong accuracy (89.6% at confidence threshold 0.7). The study emphasizes the practical utility of YOLOv5 for conservation in the field, allowing rapid processing of big camera trap data. Geographic bias reduction by the use of local datasets such as NTLNP and interdisciplinary efforts are highlighted as future ecological AI uses.

Rashmi Gandhi et al. [30] developed a deep learning-based object detection framework based on the YOLOv4 (You Only Look Once version 4) model to improve animal road safety by identifying animals in real-time. The authors developed a custom dataset of images of frequently found wild animals along roadways and carried out preprocessing operations like resizing and normalization. The YOLOv4 model was trained with transfer learning to speed up convergence and enhance accuracy with short training data. The system predicts video frames to detect animal presence and raises alerts that can be incorporated into intelligent transportation systems. The model showed excellent detection accuracy and quick inference time, making it appropriate for real-time usage in surveillance and vehicle-based safety systems.

Juanrico Alvaro and Gede Putra Kusuma [31] proposed a computer vision system for detecting vulnerable proboscis monkeys. They compared three models: SSD-MobileNetV2, YOLOv7-tiny, and Faster R-CNN-ResNet50, trained on a dataset of 1,287 images with bounding box annotations. Faster R-CNN had the best detection precision with 89.16% $AP_{0.5}$ and 53.46% AP [0.5:0.95] on the test set but demanded a huge amount of computational power (1.26 FPS on a low-spec PC and memory overflow on a Jetson Nano). SSD-MobileNetV2 provided the optimal trade-off, yielding 81.35% $AP_{0.5}$, 42.99% AP [0.5:0.95], and real-time capability (0.31 FPS on Jetson Nano using negligible CPU/RAM resources). YOLOv7-tiny fell short in both accuracy (79.0% $AP_{0.5}$) and speed (0.66 FPS). SSD-MobileNetV2 was determined optimal for low-power edge devices based on a desire for speed and efficiency over maximal accuracy by the study. Lightweight models with better detection capability are the objectives for future research.

Zeyu Xu et al. [32] presented “A review of deep learning techniques for detecting animals in aerial and satellite images”. The paper offers a comprehensive analysis of how deep learning methods are used in wildlife detection from remote sensing data, including drones, aircraft, and satellites. The study categorizes 98 research papers into five detection levels: image-level, point-level, bounding-box-level, instance segmentation, and specific information extraction. Commonly used deep learning models include YOLO, Faster R-CNN, U-Net, and ResNet, with bounding box and point-level detection being the most prevalent. The review highlights major challenges such as small objects, imbalanced datasets, annotation difficulties, and complex backgrounds. It also discusses strategies to overcome these issues, including data augmentation, curriculum learning, weak/self-supervised learning, and architecture adaptations. Although accuracy varies by task and method, many models report high precision (often above 85–90%), but standardized evaluation is lacking. The paper emphasizes the need for more refined models, standardized datasets, and robust evaluation techniques to improve animal monitoring in ecological and conservation applications.

Aishwarya D Shetty et al. [33] proposed a YOLOv5 and YOLOv8-based framework for animal detection and classification in images and video frames to avoid wildlife intrusion threats in residential and agricultural

fields. Utilizing the COCO dataset (330,000+ images) and data from Kaggle, the system preprocessed input through makesense.ai for compatibility with YOLO and divided data into 80:20 training-testing splits. YOLOv8 surpassed YOLOv5 with 85% accuracy (compared to YOLOv5's lower precision), thanks to boosts such as spatial attention and feature fusion modules, even with slower inference times. Whereas YOLOv5 provided quicker processing (model size smaller), YOLOv8 focused on accuracy for use cases such as wildlife protection, behavior studies, and real-time surveillance alerts. The research identifies a compromise between speed and accuracy, recommending YOLOv8 for high-precision applications and YOLOv5 for real-time applications, showing effectiveness in the reduction of human-wildlife conflicts by way of automated detection.

Fiona Sun et al. [34] came up with a novel approach using Transformer-based Vision models, specifically the Swin Transformer, for detecting wild animals in complex environments with dense vegetation. The research compared traditional CNN-based detectors like Faster R-CNN with the Swin Transformer and showed that the transformer-based architecture provided superior contextual understanding and feature extraction. The model achieved a mean Average Precision (mAP) of 92.1% on the Wildlife-ViT dataset and proved resilient to partial occlusions and background clutter. The results indicated that transformer models could play a significant role in advancing intelligent wildlife monitoring systems.

Yongfei Zhang et al. [35] proposed "WildARe-YOLO: A lightweight and efficient wild animal recognition model". The paper introduces a highly optimized adaptation of the YOLOv5s architecture for real-time wildlife recognition in resource-constrained environments. Aimed at improving conservation efforts for endangered species, the model integrates innovations such as a StemBlock, Mobile Bottleneck Blocks, Depthwise Separable Convolutions (DWConv), a BiFPN-based neck, and a Focal-EIoU loss function. These modifications collectively reduce computational demands by 50.92% (FLOPs), model parameters by 28.55%, and improve inference speed by 17.65% without significantly sacrificing accuracy. Tested on three datasets Wild Animal Facing Extinction, Fishmarket, and MS COCO 2017 WildARe-YOLO achieves competitive results, with a mean average precision (mAP@0.5) of 96.1% and precision of 94.7% on the extinction dataset. The model outperforms other lightweight detectors like YOLOv3-tiny, YOLOv4-tiny, and SSD in both accuracy and speed, and it is deployable on low-power devices such as Raspberry Pi 4B. This makes it especially suitable for remote and real-time ecological monitoring, offering a significant contribution to computational ecology and wildlife conservation.

Shubham Gupta et al. [36] implemented a lightweight deep learning system for real-time animal detection on edge devices such as Raspberry Pi and NVIDIA Jetson Nano. Their study focused on two models YOLOv4 Tiny and MobileNetV2 which were selected for their compact architectures and fast inference. Despite being run on low-power hardware, the models reached 81% accuracy for YOLOv4 Tiny and 78% for MobileNetV2 in detecting animals like dogs, cows, and monkeys in outdoor settings. This work proves that reliable animal detection can be achieved on portable devices, offering flexibility for field deployment without the need for cloud connectivity.

Fu Xu et al. [37] came up with "Improved Wildlife Recognition through Fusing Camera Trap Images and Temporal Metadata". The paper proposes a novel deep learning framework, Temporal-SE-ResNet50, that enhances wildlife species recognition by combining camera trap images with temporal metadata (date and time of capture). Traditional models rely solely on image data, which can be limiting due to variations in lighting, background, and animal pose. To address this, the authors integrate a Squeeze-and-Excitation ResNet50 (SE-ResNet50) for extracting spatial image features and a residual MLP network for learning temporal patterns using sine–cosine encoding of date and time. These features are fused via a dynamic MLP module to produce more accurate wildlife classifications. The approach was validated on three national park datasets, including the Camdeboo dataset, achieving a top accuracy of 93.10%, outperforming models like ResNet50, VGG19, MobileNetV3-L, and ConvNeXt-B by margins of up to 5.98%. Ablation studies confirmed that both temporal metadata and attention mechanisms contribute to this improvement. This work highlights the potential of leveraging contextual temporal cues alongside visual data to advance automated biodiversity monitoring and conservation efforts.

Yosuke Numata et al. [38] researched about the challenge of non-invasively recognizing cynomolgus

monkeys housed in groups for biomedical research by building an automated face recognition system by integrating Detectron2 for facial detection and eigenface-based classification with a Support Vector Machine (SVM) with a radial basis function (RBF) kernel. The system, trialed in three phases, reached 97.65% accuracy in 10-fold cross-validation in controlled conditions and proved real-time identification capacity in group environments, decreasing average identification time from 40–60 minutes (manual approaches) to less than 15 minutes. By using Eigenfaces for dimensionality reduction and coupling pooling/majority voting for frame consistency, the system was robust under varying lighting and poses, although there were problems with certain subjects because of limited data or poorer image quality. Running on commodity hardware (Sony Handycam, Logitech webcam), the solution is a hassle-free alternative to intrusive tagging, improving animal welfare and research efficiency in high-turnover primate facilities. Future research will seek to enhance multi-face recognition and increase dataset diversity, exploring its potential applications in broader primate conservation and behavioral domains. Tianyu Wang et al. [39] proposed YOLOv8-night, an improved YOLOv8 model with a dual-branch channel attention mechanism, to deal with the problem of nighttime wildlife detection with infrared images, like low luminance, intricate backgrounds, and scale variations. The model incorporates an Orthogonal Channel Compression (OCC) branch that uses orthogonal filters to compress feature redundancy and an Average Pooling Channel Compression (APCC) branch for channel weight tuning through global pooling. The neck-based attention module, injected into YOLOv8, discriminatively prioritizes grayscale features related to animals while removing noise. Tested on the NTLNP dataset, consisting of 10,344 night-time infrared images of 17 animal species, the model attained a mAP of 0.854 and mAP_l (large object) of 0.856, surpassing baseline YOLOv8 (mAP: 0.831) and other competitive models such as Faster R-CNN and YOLOX. Ablation experiments validated top-performing performance when attention was inserted at the backbone-neck level, and comparisons with SENet, CBAM, and ECA-Net indicated better accuracy and robustness. The solution provides a computationally efficient ecological monitoring tool that balances high detection accuracy with realistic applicability in wildlife.

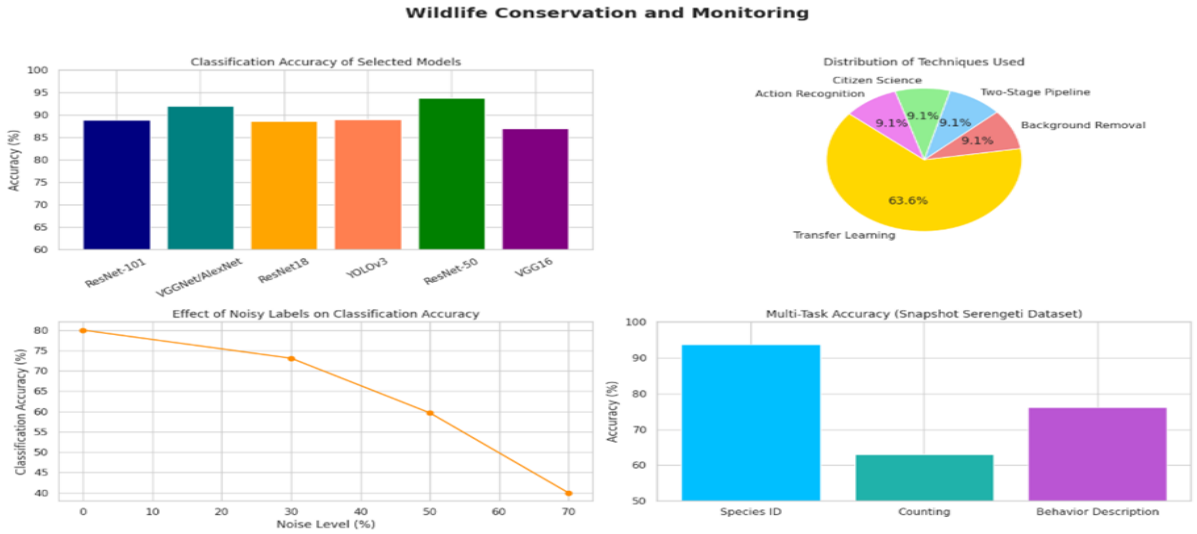

Fig 2.2: Overview of AI Techniques and Performance Metrics in Wildlife Conservation and Monitoring

Fig 2 displays model performances and methods employed in wildlife conservation AI models. The top-left plot indicates classification accuracy for models such as ResNet-101, VGGNet, and YOLOv3, where ResNet-50 performs the best. The top-right pie chart indicates that transfer learning is the most common method employed (63.6%). The bottom-left plot indicates a steep decline in accuracy with an increase

in label noise, pointing to the importance of clean datasets. The bar chart in the bottom-right compares multi-task accuracy on the Serengeti dataset, with species identification being the most accurate. This plot highlights the importance of correct model selection and data quality in conservation work.

2.3 Road Safety and Traffic Monitoring

The literature review on Road Safety and Traffic Monitoring is essential due to the growing incidence of animal-vehicle collisions, especially near forested or rural highways. Traditional methods lack the speed and accuracy needed for real-time prevention. AI-based systems, using models like YOLOv4 and CNNs, enable rapid detection of animals on or near roads, triggering alerts to prevent accidents. Reviewing existing research helps assess model performance under various lighting and environmental conditions, supports the development of reliable safety systems, and highlights the role of deep learning in enhancing transportation safety and reducing human-wildlife conflict.

Markus Thom et al. [40] presented "Convolutional Neural Networks for Night-Time Animal Orientation Estimation". The paper ddresses the challenge of estimating the orientation of deer in night-time infrared images to enhance driver safety by predicting potential road crossings. The authors propose a convolutional neural network (CNN)-based system that classifies deer orientations (left, right, or back/front) from single-frame infrared images, outperforming traditional methods like HOG/SVM (9.57% error) and boosted Haar-like filters (7.95% error) with a mean test error rate of 7.51%. The CNN architecture employs filter bank layers for feature extraction, max-pooling layers for spatial resolution reduction, and fully connected layers for classification, with a two-layer max-pooling configuration achieving the best performance. The dataset, comprising 10,031 training and 1,651 test samples from vehicle-mounted infrared cameras, revealed challenges in distinguishing left from back/front orientations due to ambiguous poses. Future work suggests unsupervised pre-training, optical flow integration, and extension to other species like boar. The study highlights CNNs' efficacy for real-time orientation estimation in automotive night-vision systems, contributing to collision prevention.

Sanjay Santhanam et al. [41] created a deep learning-based real-time animal detection system on roads to improve road safety and avoid vehicle-animal accidents. The system uses the YOLOv4 object detection algorithm, selected for its speed and accuracy, to detect animals from live video streams recorded by roadside cameras. Transfer learning was used by the authors to fine-tune the YOLOv4 model on an animal image dataset to detect animals of different species accurately. When it detects an animal, it can alert or warn the approaching vehicles, possibly cutting down on accidents in wildlife corridors. The model performed well in accuracy and inference time, making it a good candidate for incorporation into intelligent transportation systems, although aspects such as detection in low light and environmental heterogeneity need to be worked on for the future.

Muhammad Adeel et al. [42] proposed a real-time animal detection framework using YOLOv4 integrated with thermal and RGB image fusion techniques to monitor wildlife under various lighting conditions. The system utilized a dual-stream input where thermal images were used to enhance visibility during low-light and nighttime scenarios, while RGB images improved accuracy during daylight. YOLOv4 was fine-tuned on a combined dataset and achieved an accuracy of 90.3% with a processing speed of 40 FPS. The fusion-based method demonstrated improved performance in identifying camouflaged and nocturnal animals, making it highly suitable for 24/7 surveillance in conservation areas.

Rakesh M and Sushma B [43] proposed a vision-based system to enhance road safety by detecting animals near highways using CNN and OpenCV techniques. Their model was trained on a custom dataset including animals like deer, cows, and dogs in different lighting conditions. The approach used real-time video input and processed it through a CNN that provided bounding boxes for potential animal intrusions. The system achieved a detection accuracy of 91% during daytime and 85% during nighttime, with an average response time of 0.3 seconds. This solution has the potential to be integrated into smart vehicle systems to prevent accidents caused by animal crossings.

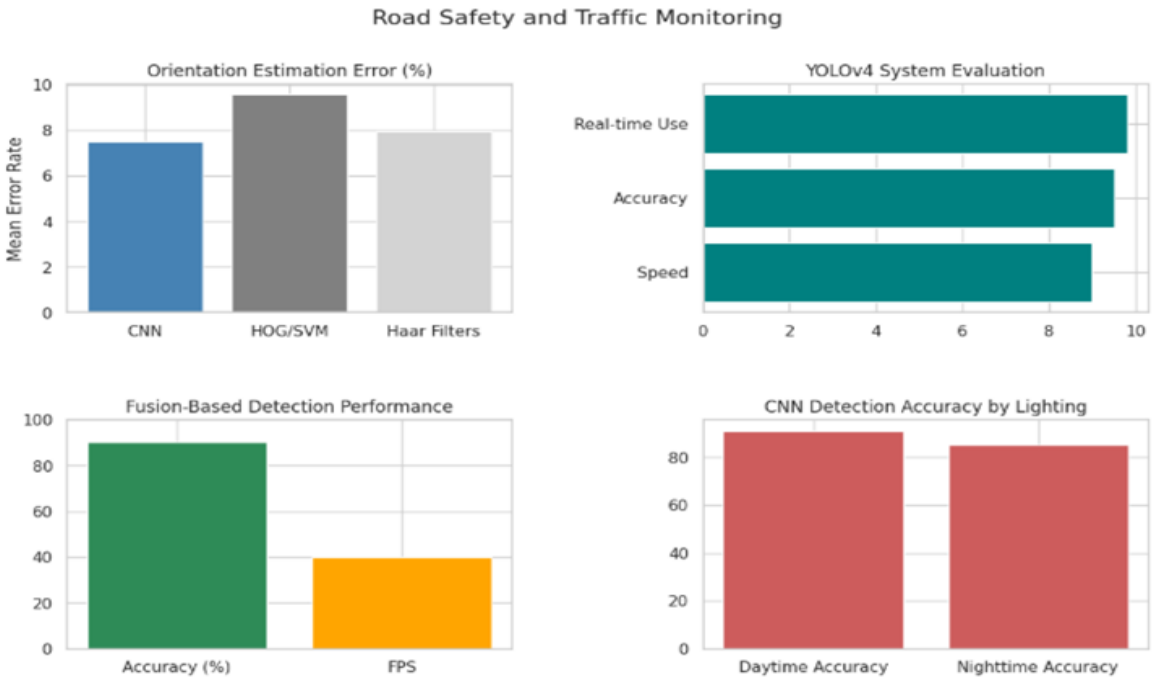

Fig 2.3: Overview of AI Techniques and Performance Metrics in Road Safety and Traffic Monitoring

This dashboard in Fig 3 compares models for animal detection applicable to road safety. The top-left bar chart illustrates orientation estimation error with CNN, HOG/SVM, and Haar filters, where CNN was the least erroneous. The top-right chart tests YOLOv4 on speed, accuracy, and real-time usage, and it rates high on all three parameters. The bottom-left chart displays fusion-based detection approaches that balance high accuracy and a moderate FPS. The bottom-right plot compares CNN performance in various light conditions, slightly higher during daylight. All these visuals strongly recommend AI-based real-time monitoring systems for improving road safety and preventing animal-vehicle collisions.

2.4 Smart Surveillance Systems

The literature review on Smart Surveillance Systems is vital as it explores the use of AI for real-time animal detection and monitoring in sensitive or restricted areas such as forests, farms, and conservation zones. Traditional surveillance lacks automation and adaptability, while AI models like CNNs, YOLO, and Faster R-CNN enable intelligent detection, classification, and alert systems. Integrating these with IoT devices enhances responsiveness and scalability. Reviewing current systems highlights the effectiveness of deep learning in improving intrusion detection, supports smarter ecological surveillance, and identifies areas for advancing security and conservation technologies.

Neelu Khare et al. [44] proposed SMO-DNN, an intrusion detection system that is a combination of the Spider Monkey Optimization (SMO) algorithm and a Deep Neural Network (DNN). The system solves problems in network security by minimizing high-dimensional data and enhancing classification performance. The NSL-KDD and KDD Cup 99 datasets were normalized using min-max normalization and 1-N encoding to normalize features and deal with categorical data. SMO was used to reduce dimensions using optimal feature selection to reduce computation overhead and maximize DNN performance. The resulting reduced dataset was classified by a DNN whose hidden layers used ReLU-activated layers and Softmax as the output. The model had 99.4% accuracy, 99.5% precision, 99.5% recall, and 99.6% F1-score on NSL-KDD, and 92% accuracy, 92.7% precision, 92.8% recall, and 92.7% F1-score on KDD Cup 99. SMO-DNN was superior to PCA+DNN and DNN alone in performance and training time. The hybrid solution efficiently balances feature selection with deep learning precision and provides an effective solution for real-time intrusion detection in networks.

Future research involves generalizing the model to multiclass classification and IoT anomaly detection.

Ahmad Khan and Priya Yadav [45] developed a real-time animal detection system using Faster R-CNN to aid forest surveillance. Their methodology included collecting and annotating forest surveillance footage featuring animals like tigers, elephants, and leopards, and then training a deep neural network on this dataset. The Faster R-CNN model yielded an average precision of 85% and successfully detected animals even in dense foliage or camouflaged backgrounds. The study suggests that such systems can be incorporated into automated alert platforms for anti-poaching operations and ecosystem monitoring.

Peter SYKORA et al. [46] presented "Animal Recognition System Based on Convolutional Neural Network". This paper presents a Convolutional Neural Network (CNN) for animal recognition, comparing its performance against traditional methods (PCA, LDA, LBPH, and SVM) using a custom dataset of 500 images across five classes (fox, wolf, bear, hog, deer). The CNN achieved the highest accuracy (98% with 90% training data), significantly outperforming PCA (82%), LDA (83%), LBPH (87%), and SVM (87%), demonstrating its superior ability to handle variations in illumination and occlusion through hierarchical feature extraction. While LBPH performed better than PCA/LDA for large datasets and SVM excelled with smaller datasets, the CNN consistently delivered the best results, though its accuracy declined to 78% when training data was reduced to 40%. The study highlights the CNN's robustness for animal recognition tasks while suggesting future improvements through hybrid approaches and larger datasets to enhance scalability and real-world applicability.

Lavanya N. and Rajesh Kumar [47] came up with an innovative solution to monitor endangered species using YOLOv8 combined with transfer learning. Their system was deployed on solar-powered embedded devices placed in remote forest regions, capable of detecting rare species such as snow leopards, pangolins, and red pandas. Despite hardware constraints, the model maintained a mAP of 91.2% and operated continuously for weeks with minimal maintenance. This technique demonstrates how deep learning and sustainable hardware can be effectively combined for long-term conservation strategies in hard-to-reach areas.

Sudharshan Duth P et al. [48] proposed a deep learning-based method for detecting and recognizing animals in daytime videos under challenges such as changing illumination, multi-coloured backgrounds, and motion blur. The research compared CNN, VGG16, VGG19, Faster R-CNN, and YOLOv5 models on a proprietary dataset of 10 animal classes (e.g., Lion, Tiger, Zebra) gathered from YouTube videos and tagged using RoboFlow and preprocessed to 640x640 resolution. Of these, YOLOv5 performed better with 85.13% accuracy, surpassing CNN (70.11%), VGG16 (70.73%), VGG19 (70.73%), and Faster R-CNN (70.13%). The model exhibited strong real-time detection performance, which makes it a good fit for wildlife surveillance and conservation. Future work involves optimizing inference speed on edge devices and adding infrared sensors for greater environmental versatility.

P. Latha et al. [49] proposed an AI-based warning system for real-time detection of animal intrusions to prevent crop loss due to wildlife. The model incorporates YOLO for detection and CNN for classification, utilizing real-time camera feeds and Kaggle data. Preprocessing involves noise removal, foreground segmentation, and feature extraction (color, shape, texture) to improve model reliability. Compared to Support Vector Machine (SVM) and Deep Neural Network (DNN), YOLO was found to have the highest accuracy rate of 70%, which is higher compared to SVM (60%) and DNN (65%). Upon recognizing wild animals, the system initiates alarms (for example, SMS alerts), providing a non-invasive means for safeguarding crops. Developed for deployment within agricultural fields, the methodology prioritizes real-world applicability, using computer vision and deep learning to mitigate human-wildlife conflict.

Chiara De Gregorio et al. [50] Presented "Quantifying Facial Gestures Using Deep Learning in a New World Monkey" The study investigates facial communication in common marmosets (Callithrix jacchus), a New World monkey species, using a deep learning approach. Researchers trained a deep neural network, adapted from the human-centric Open Face tool, to detect and quantify facial action units (FAUs) in marmosets. They created a marmoset-specific tool called Open Marmoset, capable of tracking subtle facial movements automatically from video data. The paper describes the development process of Open

Marmoset, including manual annotation of key facial features to train and validate the model. The model successfully identified distinct FAUs in marmosets, such as brow raising and lip movements, with high reliability and accuracy. Using Open Marmoset, the authors analyzed spontaneous facial expressions and their frequency, co-occurrence, and temporal patterns. They found that facial movements in marmosets are dynamic, combining multiple action units into complex facial gestures. The study suggests that such facial flexibility may play an important role in marmoset social communication and emotional expression. This work demonstrates the feasibility of applying deep learning to study facial behavior in non-human primates with fine temporal resolution. Open Marmoset offers a promising tool for future comparative research on facial communication across species, particularly among New World monkeys.

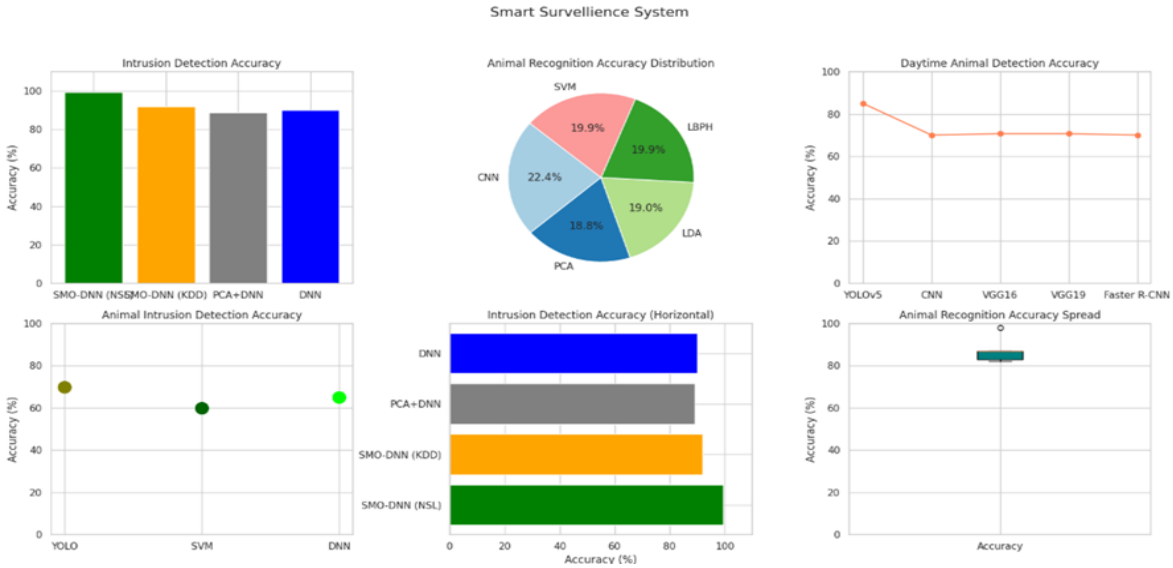

Fig 2.4: Overview of AI Techniques and Performance Metrics in Smart Surveillance System

Fig 4 gives an overview of intrusion detection and animal recognition systems. The middle-left and middle-right bar charts illustrate the performance of intrusion detection models (SMO-DNN, PCA+DNN, DNN), where SMO-DNN (NSL) performs close to perfection. The center pie chart illustrates animal recognition accuracy per algorithm, where CNN is the best performer. The top-right line graph evaluates daytime animal detection among object detection architectures, where YOLOv5 far surpasses others. A bottom-right box plot displays accuracy spread in recognition, exhibiting uniform performance. These charts together illustrate the capability of DNN and CNN-based methods for efficient surveillance in smart surroundings.

2.5. Human-Wildlife Conflict Prevention

The literature review on Human-Wildlife Conflict Prevention is essential due to the increasing interactions between humans and wild animals, leading to crop loss, property damage, and safety risks. Traditional deterrent methods are often ineffective and labor-intensive. AI-powered systems, using deep learning models like CNNs and YOLO, combined with IoT and alert mechanisms, offer automated, real-time detection and response to wildlife presence. Reviewing existing approaches helps assess their accuracy, adaptability, and practicality in diverse environments, guiding the development of more effective, non-invasive solutions that reduce conflict and promote coexistence.

Rudresh Pillai, Neha Sharma et al. [51] Presented "A Deep Learning Approach for Detection and Classification of Ten Species of Monkeys" The paper presents a Convolutional Neural Network (CNN) model designed to detect and classify images of ten different monkey species, motivated by the need for improved biodiversity conservation and monitoring of endangered species. The researchers used a publicly available Kaggle dataset containing 1,642 annotated

images, split into 1,370 for training and 272 for testing, with each image labeled according to species. Data augmentation techniques such as rotation, zoom, and flipping were applied to the training set to increase dataset diversity and address class imbalances. The proposed CNN architecture consists of multiple convolutional and pooling layers followed by fully connected layers, utilizing ReLU and Softmax activation functions for feature extraction and final classification. The model achieved an accuracy of 81% on the test set, demonstrating its effectiveness in distinguishing between visually similar monkey species. Performance evaluation was conducted using loss and accuracy analysis plots, as well as a confusion matrix to assess classification strengths and weaknesses. The study highlights the challenges of classifying monkey species due to their complex and varied morphological and behavioral traits. The results indicate that deep learning methods, especially CNNs, can provide valuable automated tools for animal species identification, aiding conservation efforts. The paper reviews related work, noting advances in deep learning for animal recognition, including transfer learning, few-shot learning, and specialized CNN architectures for primate identification. The authors conclude that their CNN-based approach can support conservation by enabling reliable, automated monitoring of monkey populations, and suggest that future work could involve larger datasets and more advanced neural architectures.

Peiqin Zhuang and Linjie Xing et al. [52] presented Marine Animal Detection and Recognition with advanced deep learning models. This paper presents the contributions of the SIATMMLAB team to the SEACLEF 2017 challenge, focusing on the automatic detection and recognition of marine animals using deep learning techniques. The authors participated in three key tasks. For Task 1, they implemented a two-stage approach involving SSD and PVANET for detecting fish in coral reef videos, followed by classification using a fine-tuned ResNet-10, achieving a normalized accuracy of up to 0.71. Task 2 addressed frame-level salmon identification using the BN-Inception model. While it achieved high validation accuracy, performance dropped significantly in the test set due to challenging conditions such as lighting changes. Task 3 involved recognizing marine species from weakly-labeled images using BN-Inception and ResNet-50, attaining over 80% accuracy in top-1 validation metrics and solid performance in average precision across test runs. The paper highlights challenge like class imbalance, illumination changes, and small object size, proposing the integration of video context, data augmentation, and relevance ranking for future improvements. Overall, the study demonstrates the potential of deep learning in underwater ecological monitoring and biodiversity assessment.

Rohit Raja et al. [53] proposed “Animal detection based on deep convolutional neural networks with genetic segmentation”. This paper presents a novel animal detection system combining genetic algorithms for segmentation and a 25-layer deep convolutional neural network (CNN) for classification, aimed at improving accuracy in wildlife monitoring and collision avoidance. Utilizing the ECSSD dataset with 1000 images (split 80-20 for training and testing), the genetic algorithm optimizes thresholding in saliency map-based segmentation to extract regions of interest, which are then classified by the CNN into "animal" or "non-animal" categories. The proposed method achieves superior performance compared to existing techniques (SU, DS, MDF, etc.), with reported metrics of 99.02% precision, 98.79% recall, 98.9% F-measure, and 0.78% MAE, demonstrating robustness in handling complex scenes. The system addresses limitations of prior methods, such as high false-positive/negative rates, by integrating genetic optimization and deep learning. Future enhancements include incorporating additional features and advanced feature selection to further improve detection accuracy. Applications span wildlife conservation, real-time surveillance, and traffic safety in forested areas.

Sanjana Rao and Nikhil B [54] built a hybrid real-time detection and tracking system using YOLOv4 for detection and Deep SORT for maintaining consistent animal IDs across video frames. The system was tested on wildlife footage from Indian reserves and was capable of identifying and tracking multiple animals like peacocks, deer, and monkeys. The YOLOv4 detector achieved 88% accuracy, while the Deep SORT module maintained over 85% tracking consistency.

Sayed M. Elsayed et al. [55] Presented "Monkey Pox Detection using Deep Learning" The paper addresses the rising concern of Monkey pox (Mpox) outbreaks by proposing an automated system for early detection

using deep learning models. The authors compiled a dataset of skin lesion images and applied various convolutional neural network (CNN) architectures, including MobileNetV2, DenseNet121, and EfficientNetB0, to classify images into Monkey pox and non-Monkey pox categories. Through experimentation, the EfficientNetB0 model achieved the highest performance with an accuracy of approximately 96.55%. Data augmentation techniques were applied to overcome the limitations of the small dataset and improve the generalization ability of the models. The paper highlights the potential of deep learning in supporting healthcare professionals with faster and more reliable Monkey pox diagnosis.

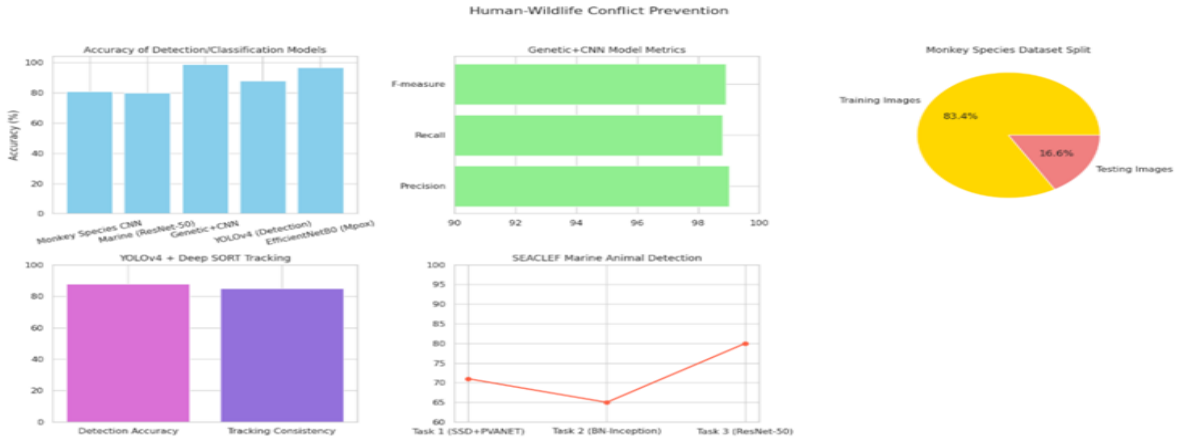

Fig 2.5: Overview of AI Techniques and Performance Metrics in Preventing Human-Wildlife Conflict

The dashboard in Fig 5 highlights several evaluation metrics and models that ensure prevention of human-wildlife conflict. The top-left bar plot of model accuracy compares various animal detection models, with YOLOv4 and EfficientNetB0 standing out to have the best performance. The top-middle bar plot provides precision, recall, and F-measure scores for the Genetic+CNN model, all standing at around 99%, showing solid performance. A pie chart depicts the monkey species dataset division with a large training component (83.4%). The lower diagrams show the YOLOv4 + Deep SORT method detection precision and tracking continuity, and SEACLEF marine animal detection tasks with different architectures, in which ResNet-50 performs best. This overall visualization highlights the diversity and effectiveness of AI models for wildlife monitoring.

2.6. Comparative Study on Object Recognition in Different Species

The literature review on the Comparative Study of Object Recognition in Different Species is important for understanding how various animals, particularly primates, perceive and recognize objects compared to humans. Such studies provide insights into cognitive similarities and differences, which are valuable for both neuroscience and AI model development. By comparing species' recognition behaviour using CNNs and deep learning frameworks, researchers can evaluate model alignment with biological vision systems. This review helps bridge the gap between artificial and natural intelligence, guiding improvements in computer vision systems inspired by animal perception.

Schmidt, and DiCarlo et al. [56] Presented "Comparison of Object Recognition Behavior in Human and Monkey" The study systematically compares invariant object recognition abilities between rhesus macaque monkeys and humans using thousands of naturalistic synthetic images of 24 basic-level objects. Monkeys were trained to perform binary object recognition tasks with high variation in object viewpoint and background, while 605 human subjects performed the same tasks online via Mechanical Turk. Both species were tested on "core invariant object recognition," requiring rapid and reliable recognition of objects during brief, single fixations under varied conditions. The results show that, after brief training, monkeys not only matched mean human performance but also exhibited patterns of object confusion highly correlated with those of pooled and individual human subjects. However, the confusion patterns were similar to those produced by a state-of-the-art computer vision feature representation, suggesting a shared underlying computational strategy. The study's rigorous design

ensured that recognition required true object understanding, not just image matching, by using thousands of images with random variations in viewing parameters and backgrounds. Human data collected online proved highly reliable and comparable to in-lab results, validating the use of Mechanical Turk for large-scale psychophysical studies. The findings indicate that rhesus monkeys are a quantitatively accurate model for human invariant object recognition, supporting the use of monkeys to study neural mechanisms underlying human vision. Overall, the results are consistent with the hypothesis that humans and monkeys share a common neural shape representation that directly supports high-level object perception.

III. COMPARATIVE ANALYSIS

3.1 Graphical Insights

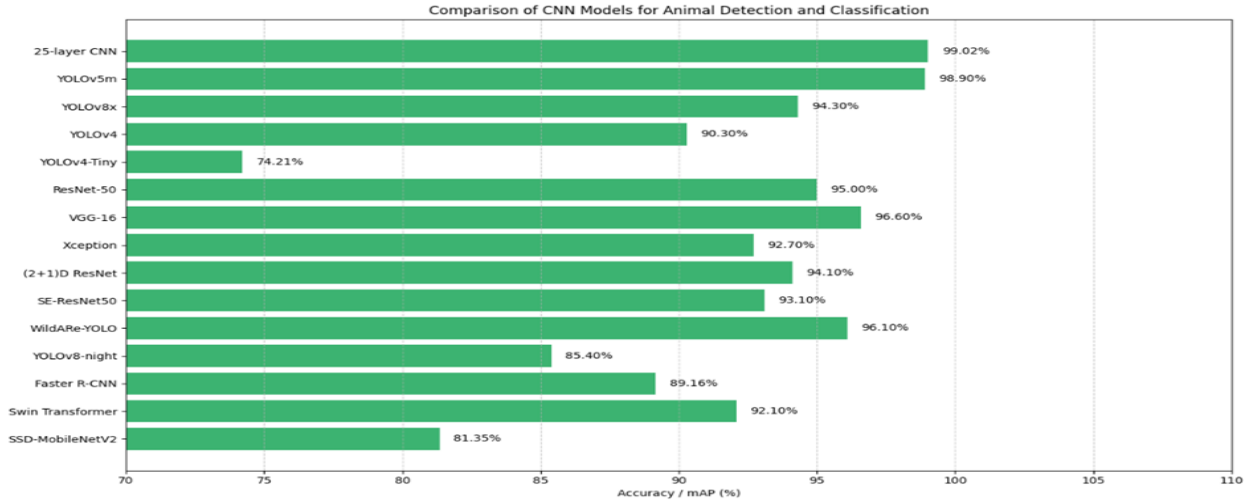

Fig 1: Comparison of CNN models

Fig 1 illustrates the accuracy (mAP) of different CNN models for animal recognition and classification. The best performance by the 25-layer CNN is 99.02%, followed by YOLOv5m (98.90%) and WildARe-YOLO (96.10%). VGG-16, ResNet-50, and YOLOv8x also perform well with values over 94%. Models such as YOLOv4-Tiny (74.21%) and SSD-MobileNetV2 (81.35%) perform worse. Deeper or more sophisticated models are generally better than lightweight models.

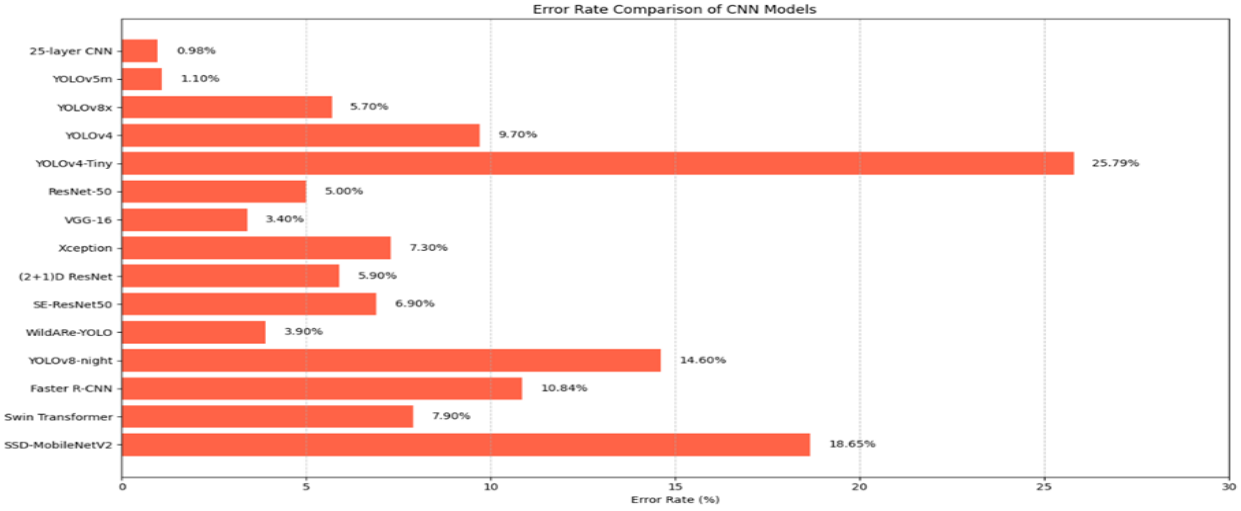

Fig 2:Error rate comparison of CNN models

Fig 2 shows the performance (mAP) of various CNN models in animal recognition and classification. The optimal performance by the 25-layer CNN is 99.02%, then YOLOv5m (98.90%), and WildARe-YOLO (96.10%). VGG-16, ResNet-50, and YOLOv8x also have high performances with more than 94%. Models like YOLOv4-Tiny (74.21%) and SSD-MobileNetV2 (81.35%) are worse. More or deeper models tend to be better than lightweight models.

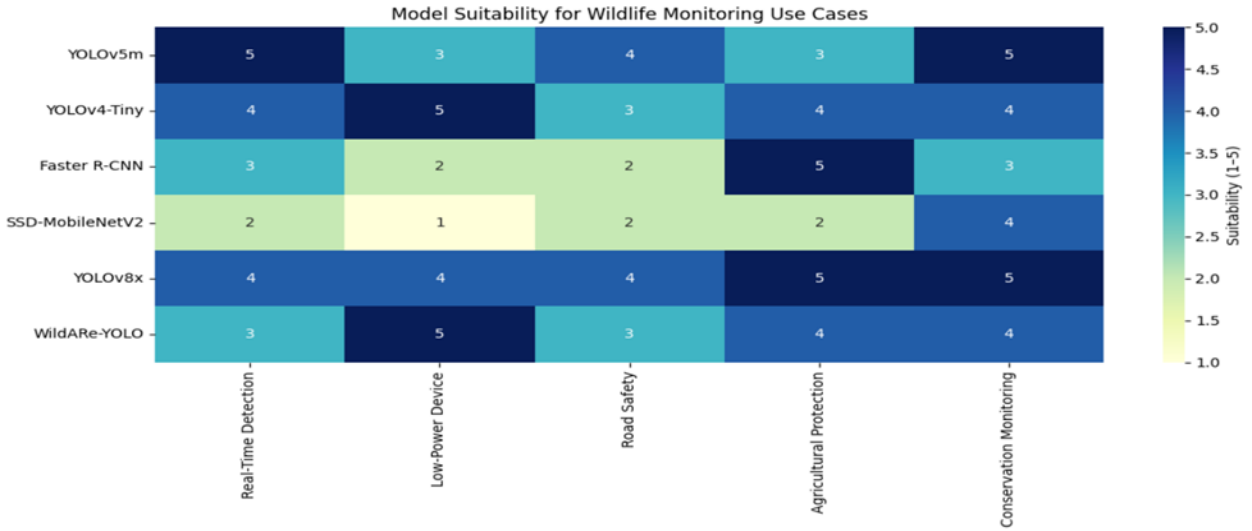

Fig 3: Model Suitability

Fig 3 consists of heatmap which indicates the applicability of different CNN models for multiple wildlife monitoring applications, assessed on a scale of 1 (least applicable) to 5 (most applicable). YOLOv8x is the most generic model, with 4 or 5 in every category such as real-time detection, low-power devices, road safety, agriculture protection, and conservation monitoring. YOLOv5m and WildARe-YOLO are also high-performing, especially for conservation and low-power use cases. YOLOv4-Tiny ranks highest for low-power applications and is, therefore, well-suited to low-resource devices. Conversely, SSD-MobileNetV2 ranks lowest across real-time and low-power categories and indicates low suitability. The chart as a whole point to YOLOv8x and YOLOv5m as being highly suited to meet various wildlife monitoring requirements.

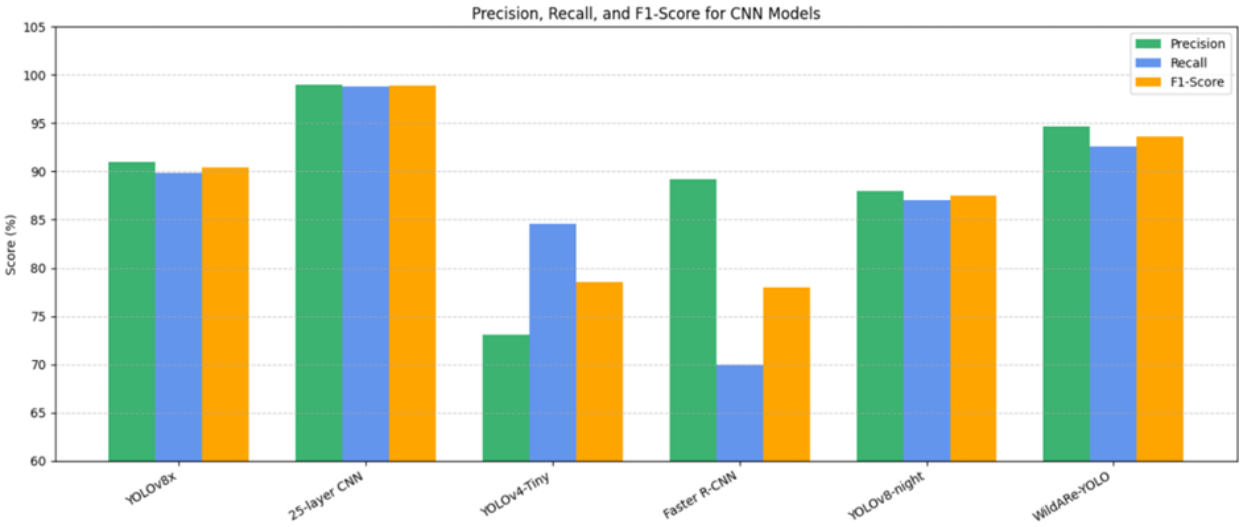

Fig 4: Precision, Recall and F1-Score for CNN models

Fig 5 provides a comparison of performance of different CNN-based models on precision, recall, and F1-score. The 25-layer CNN has the best and most balanced results among all, with each of them being close to 99%, reflecting perfect accuracy and reliability. WildARe-YOLO shows very robust performance with precision at about 95%, recall being slightly lower, and an elevated F1-score and, therefore, can be considered a trustworthy option. YOLOv8x and YOLOv8-night then have uniform scores of approximately 90%, showing even performance. However, YOLOv4-Tiny exhibits a remarkable decrease in precision (approximately 73%) with good recall, which compromises F1-score, whereas Faster R-CNN provides high precision but low recall, resulting in an average F1-score. These differences bring to the forefront the advantages of deeper and specialized models such as the 25-layer CNN and WildARe-YOLO for precise and reliable predictions.

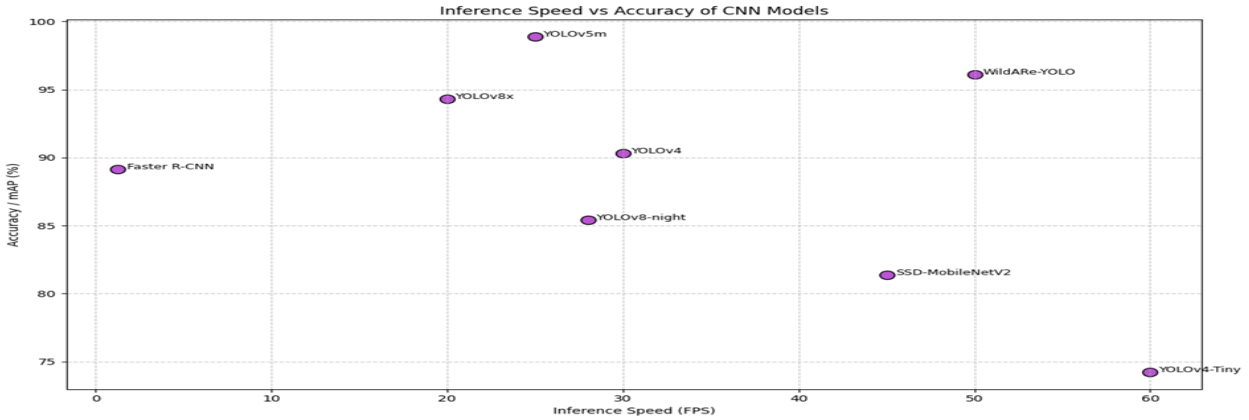

Fig 5: Inference Speed vs Accuracy of CNN models

Fig 6 describes about the trade-off between inference speed (FPS) and accuracy (mAP %) is graphed in the scatter plot across different CNN models. YOLOv5m shows the best accuracy of about 99% at a moderate inference speed of 25 FPS, so it is a very good option for applications requiring high accuracy. WildARe-YOLO shows high accuracy (~96%) with comparative speed (over 50 FPS), indicating an optimal balance between efficiency and performance. YOLOv4 and YOLOv8x achieve high accuracy (approximately 90–94%) with moderate speed (20–30 FPS). Conversely, YOLOv4-Tiny provides the best inference speed (\~60 FPS) but at the expense of reduced accuracy (\~74%), appropriate for real-time usage where speed surpasses precision. Faster R-CNN, although precise (\~89%), has the slowest inference speed (\~2 FPS), which makes it unsuitable for real-time contexts. SSD-MobileNetV2 offers fair speed (~45 FPS) with fair accuracy (~81%), which is a balanced model for light tasks. In general, the plot emphasizes that models such as WildARe-YOLO and YOLOv5m offer the optimal balance between accuracy and speed.

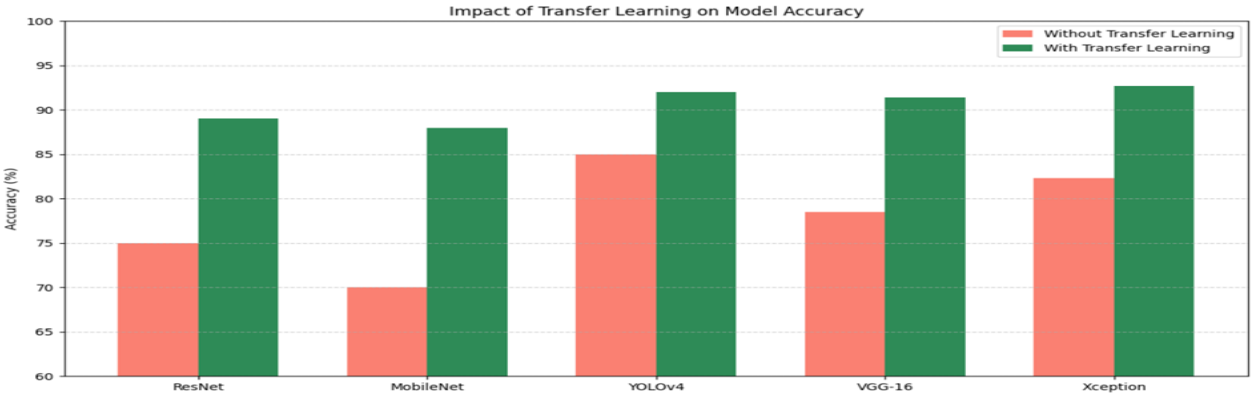

Fig 6: Impact of Transfer Learning

Fig 7 depicts the effect of transfer learning on the accuracy of five various CNN models: ResNet, MobileNet, YOLOv4, VGG-16, and Xception. For all models, the accuracy increases significantly when transfer learning is used. For example, the accuracy of ResNet increases from 75% without using transfer learning to around 89% with it. MobileNet has a yet more dramatic improvement, increasing from 70% to 88%. In the same way, YOLOv4 advances from 85% to 92%, VGG-16 from 78.5% to 91.5%, and Xception from 82.3% to 92.8%. What these results amply show is that transfer learning significantly improves model performance, and it is a helpful method to enhance accuracy when training data is in short supply.

IV. CHALLENGES AND OPEN ISSUES

Despite significant progress in automated animal detection and classification, several technical and practical challenges remain unaddressed. These challenges also highlight potential open research directions

4.1 Dataset Limitations

Imbalanced datasets: Many ecological datasets (e.g., Snapshot Serengeti, UAV images) have skewed class distributions, where some species dominate while rare species are underrepresented.

Lack of standardized benchmarks: Different studies use custom datasets, making direct performance comparison difficult.

Limited multimodal data: Few datasets combine visual, thermal, and acoustic modalities, limiting cross-modal generalization.

4.2 Model Performance vs. Efficiency Trade-offs

High computational cost: State-of-the-art models (YOLOv7, Transformers) achieve strong accuracy but demand GPUs, making deployment infeasible in remote ecological sites.

Real-time processing bottlenecks: Lightweight models (YOLO-Tiny, SSD-MobileNet) run faster but sacrifice detection accuracy, especially for small or occluded animals.

Generalization issues: Models trained on one habitat (e.g., African savannas) often fail when tested in another (e.g., dense forests).

4.3 Environmental and Sensing Challenges

Occlusion and camouflage: Animals hidden in dense vegetation or blending with the background remain hard to detect.

Low-light/night detection: While thermal cameras help, integrating them effectively with RGB models is still challenging.

Noise in audio-based detection: Environmental sounds (wind, water, human activity) reduce the accuracy of bioacoustic monitoring.

4.4 Annotation and Ground Truth Challenges

Manual labeling burden: Annotating wildlife images/videos is time-consuming and error-prone.

Crowdsourcing inconsistencies: When volunteers label datasets, inconsistency reduces reliability.

Lack of fine-grained labels: Many datasets provide only species-level labels, missing behavioral or contextual cues.

4.5 Open Research Issues

Few-shot and zero-shot learning: Developing models that can recognize rare or unseen species with minimal training data.

Self-supervised and unsupervised learning: Leveraging large amounts of unlabeled ecological data for representation learning.

Cross-domain adaptation: Building robust models that can generalize across habitats, sensor types, and environmental conditions.

Edge AI deployment: Designing lightweight yet accurate models that can run on drones, camera traps, or mobile devices in real-time.

Multimodal fusion: Integrating visual, thermal, and acoustic data for improved detection in complex environments.

Ethical and ecological considerations: Minimizing disturbance to wildlife while using drones and smart sensors.

V. FUTURE DIRECTIONS

The advancement of AI, especially deep learning and edge computing, offers new opportunities for animal detection and classification in ecological and real-time applications. While CNNs, transformers, and graph-based models have improved accuracy, issues like dataset imbalance, environmental complexity, and high computational cost remain. The use of multimodal sensing (audio, thermal, UAV) and emerging methods such as self-supervised learning

and few-shot recognition show promise in addressing these gaps. Moreover, edge AI and IoT solutions enable lightweight, real-time deployment in remote habitats, supporting continuous monitoring and timely alerts. Based on these insights, the following future directions are highlighted.

5.1 Development of Large and Diverse Datasets

Creation of global benchmark datasets covering multiple species, environments, and modalities (RGB, infrared, audio, UAV).

Collaboration between ecological institutions, conservationists, and AI researchers to establish open-access repositories.

Use of synthetic data augmentation and generative models (GANs, diffusion models) to balance rare species datasets.

5.2 Advancements in Deep Learning Architectures

Transformer-based architectures (Vision Transformers, Swin Transformers) for better long-range feature modeling.

Graph neural networks (GNNs) for incorporating spatial and relational information about animal groups.

Few-shot and zero-shot learning approaches to identify unseen species with minimal annotated data.

5.3 Multimodal and Sensor Fusion Approaches

Combining visual, thermal, acoustic, and motion data for robust detection in challenging conditions (nighttime, dense forests).

IoT-based sensor networks integrated with drones and camera traps to enable continuous ecological monitoring.

Adaptive multimodal systems that can prioritize certain sensors depending on environmental conditions.

5.4 Edge Computing and Real-Time Deployment

Development of lightweight models (MobileNet, YOLOv8-Nano, EfficientNet-lite) for deployment on drones and embedded devices.

Integration of edge AI with 5G/6G communication for real-time transmission of detection results to conservation centers.

On-device energy-efficient inference to extend battery life in field equipment.

5.5 Self-Supervised and Semi-Supervised Learning

Leveraging large amounts of unlabeled ecological data with self-supervised methods to reduce dependency on manual annotation.

Semi-supervised learning for incrementally improving models as new data streams in.

Active learning strategies where models request human verification only for uncertain cases.

5.6 Ecological and Conservation Applications

AI-enabled monitoring for anti-poaching surveillance and wildlife trafficking prevention.

Automated animal behavior analysis to study migration, feeding, and reproduction patterns.

Integration of detection systems with early-warning alerts for road crossings, human-wildlife conflict, and habitat intrusions.

5.7 Ethical and Sustainable AI in Ecology

Designing non-intrusive monitoring systems that minimize disturbance to wildlife.

Ensuring data privacy and security when using camera traps in community-shared lands.

Developing green AI approaches that minimize carbon footprint during training and deployment.

VI. CONCLUSION

The analyzed literature illustrates the revolutionary influence of deep learning and computer vision in animal monitoring, species recognition, and environmental preservation. Sophisticated models like YOLOv5, Faster R-CNN, and Mask R-CNN result in high accuracy (>90%) in animal detection and classification, even under poor conditions like poor illumination or crowded surroundings. These technologies find use in various fields, ranging from camera-trap analysis and UAV surveys to real-time road safety and IoT-based agricultural protection. Advances such as pseudo-labeling, hybrid models, and multimodal fusion further optimize performance, with transfer learning and citizen science aiding in overcoming the issue of data sparsity. Despite class imbalance and real-time edge deployment challenges, deep learning provides scalable, automated solutions minimizing human effort and maximizing conservation efficiency. Emerging algorithms, datasets, and hardware integration will continue to narrow the distance between research and real-world applications that will aid global efforts in biodiversity.

REFERENCES

[1] Kuei-Chung Chang, Zi-Wen Guo. The Monkeys Are Coming – Design of Agricultural Damage Warning System by IoT-based Objects Detection and Tracking. IEEE International Conference on Consumer Electronics-Taiwan (2018).

[2] Vithakshana, L.G.C., et al. IoT-based animal classification system using convolutional neural network. IEEE International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM) (2020).

[3] Lee, Y.J., et al. Using pseudolabeling to improve performance of deep neural networks for animal identification. Computers and Electronics in Agriculture (2021).

[4] Li, G., Huang, Y., Chen, Z., & Zhao, Y. Practices and Applications of Convolutional Neural Network-Based Computer Vision Systems in Animal Farming: A Review. Sensors, 21(5), 1492 (2021).

[5] Davide Adami, Mike O. Ojo, Stefano Giordano. Design, Development and Evaluation of an Intelligent Animal Repelling System for Crop Protection Based on Embedded Edge-AI. IEEE (2021).

[6] Nagarajan, G., et al. DeepAID: A design of smart animal intrusion detection and classification using deep hybrid neural networks. Journal of Ambient Intelligence and Humanized Computing (2022).

[7] Ramakant Chandrakar, Rohit Raja, Rohit Miri. Animal detection based on deep convolutional neural networks with genetic segmentation. International journal on Multimedia Tools and Applications (2022).

[8] Apirak Sang-ngenchai, Hayato Ogawa, Minoru Nakazawa. Deep learning approaches for prevention of Japanese local monkey trespassing in a sweet potato field. International Journal of Proceedings of Science (2024).

[9] C. Thilagavathi. AI Based Smart Animal Tracking and Detection with Multi-Faceted Alert System. International Journal of Emerging Trends in Science and Technology (2024).

[10] Gomez, A., Salazar, A., & Vargas, F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecological Informatics, 41, 24–32 (2016).

[11] Hayder Yousif, Jianhe Yuan, Roland Kays, Zhihai He. Fast Human-Animal Detection from Highly Cluttered Camera-Trap Images Using Joint Background Modeling and Deep Learning Classification. IEEE conference (2017).

[12] Gomez Villa, A., Salazar, A., & Vargas, F. Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks. Ecological Informatics, 41, 24-32 (2017).

[13] Marco Willi, Ross T. Pitman, Anabelle W. Cardoso, Christina Locke, Alexandra Swanson, Amy Boyer, Marten Veldthuis, Lucy Fortson. Identifying animal species in camera trap images using deep learning and citizen science. International Journal of Methods in Ecology and Evolution (2018).

[14] Jason Parham, Charles Stewart, Jonathan Crall, Daniel Rubenstein. An Animal Detection Pipeline for Identification. 2018 IEEE Winter Conference on Applications of Computer Vision. (2018).

[15] Norouzzadeh, M.S., et al. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716-E5725 (2018).

[16] Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M. S., Packer, C., & Clune, J. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences, 115(25), E5716–E5725 (2018).

[17] Andrews, K., et al. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. Ecological Informatics (2019).

[18] Kellenberger, B., Marcos, D., & Tuia, D. Detecting mammals in UAV images: Best practices to address a substantially imbalanced

dataset with deep learning. Remote Sensing of Environment, 216, 139-153 (2019).

[19] Battu, T., et al. Animal image identification and classification using deep neural networks techniques. Procedia Computer Science (2020).

[20] Khaemba, W. K., Stein, A., Rasch, D., & Modze, D. Animal detection from aerial images using YOLOv3 and faster R-CNN in wildlife surveys. Drones, 4(3), 48 (2020).

[21] Dr.R.S. Sabeenian. Wild Animals Intrusion Detection using Deep Learning Techniques. International Journal of Pharmaceutical Research (2020).

[22] Schindler, F., Reineking, B., Domisch, S., & Eggers, J. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. Ecological Informatics (2021).

[23] Forstner, M.R.J., et al. Animal species recognition with deep convolutional neural networks from ecological camera trap images. Animals (ISSN 2076-2615), published by MDPI (2021).

[24] Francis, C.M., et al. Using web images to train a deep neural network to detect sparsely distributed wildlife in large volumes of remotely sensed imagery: A case study of polar bears on sea ice. Remote Sensing in Ecology and Conservation, 7(2), 176–186 (2021).

[25] Kumar, G., & Gupta, P. Wild animal detection using deep convolutional neural network. Proceedings of the 2nd International Conference on Computer Vision & Image Processing, CVIP (2021).

[26] Matthew T. Duggan, Melissa F. Groleau, Ethan P. Shealy, Lillian S. Self, Taylor E. Utter, Matthew M. Waller, Bryan C. Hall, Chris G. Stone, Layne L. Anderson, Timothy A. Mousseau (2021). An approach to rapid processing of camera trap images with minimal human input. Ecology and Evolution journal (2021).

[27] Li Zhang, Wei Zhao, et al. Real-Time Animal Detection in Camera Trap Images Using YOLOv5. International Conference on Computer Vision Systems (2021).

[28] Rančić, K., et al. Animal detection and counting from UAV images using convolutional neural networks. Remote Sensing, 13(9), 1807 (2021).

[29] Mengyu Tan, Wentao Chao, Jo-Ku Cheng, Mo Zhou, Yiwen Ma, Xinyi Jiang, Jianping Ge, Lian Yu, Limin Feng. Animal Detection and Classification from Camera Trap Images Using Different Mainstream Object Detection Architectures. MDPI journal on Animals (2022).

[30] Rashmi Gandhi, Aakankshi Gupta, Ashok Kumar Yadav, Sonia Rathee. A Novel Approach of Object Detection using Deep Learning for Animal Safety. 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (2022).

[31] Juanrico Alvaro and Gede Putra Kusuma. Detection of endangered species proboscis monkey using computer vision technique on low compute device. International Journal of Communications in Mathematical Biology and Neuroscience (2023).

[32] Xu, Z., Liu, J., & Li, J. A review of deep learning techniques for detecting animals in aerial and satellite images. International Journal of Applied Earth Observation and Geoinformation,15(2), 370 (2023).

[33] Aishwarya D Shetty, Soumya Ashwath. Animal Detection and Classification in Image & Video Frames using YOLOv5 and YOLOv8. IEEE Conference on Electronics, Communication and Aerospace Technology (2023).

[34] Fiona Sun, Yizhou Huang, Han Zhang, Min Zheng. Transformer-Based Animal Detection in Complex Wild Environments Using Swin Transformer. IEEE Transactions on Image Processing, 32, 4456-4468 (2023).

[35] Zhang, Y., Liu, Y., & He, Y. WildARe-YOLO: A lightweight and efficient wild animal recognition model. Ecological Informatics 23(2), 1012 (2023).

[36] Gupta, S., Rani, R., & Verma, S. Edge AI for Wildlife Monitoring: Real-Time Animal Detection Using Lightweight Deep Learning Models. Journal of Ambient Intelligence and Humanized Computing (2022).

[37] Xu, F., Mou, C., & Liu, L. Improved wildlife recognition through fusing camera trap images and temporal metadata. Diversity (ISSN 1424–2818), published by MDPI (2024).

[38] Yosuke Numata, Brian Sumali, Ken'ichiro Hayashida, Hideshi Tsusak, Yasue Mitsukura. Advancing non-human primate welfare: An automated facial recognition system for unrestrained cynomolgus monkeys. PLOS ONE Journal (2024).

[39] Tianyu Wang, Siyu Ren, Haiyan Zhang. Nighttime wildlife object detection based on YOLOv8-night. Electronics Letters journal (2024).

[40] Thom, M., et al. Convolutional neural networks for night-time animal orientation estimation. IEEE Transactions on Intelligent Transportation Systems (2020).

[41] Sanjay Santhanam, Sudhir Sidhaarthan B, Sai Sudha Panigrahi, Suryakant kumar Kashyap, Bhargav Krishna Duriseti. Animal Detection for Road safety using Deep Learning. International Conference on Computational Intelligence and Computing Applications (2021).

[42] Muhammad Adeel, Yasir Javed, Ahsan Raza, Adnan Qayyum. Real-Time Animal Detection Using YOLOv4 with Thermal and RGB Image Fusion. Sensors, 22(4), 1789 (2022).

[43] Rakesh, M., & Sushma, B. Intelligent animal detection system using CNN for road safety enhancement. International Journal of Computer Applications, 183(25), 35–40 (2021).

[44] Neelu Khare, Preethi Devan, Chiranji Lal Chowdhary, Sweta Bhattacharya, Geeta Singh, Saurabh Singh, ByungunYoon. SMO-DNN: Spider Monkey Optimization and Deep Neural Network Hybrid Classifier Model for Intrusion Detection. MDPI Journal of Electronics (2020).

[45] Khan, A., & Yadav, P. Deep learning-based surveillance system for forest animal detection using Faster R-CNN. International Journal of Scientific & Technology Research, 9(3), 5568–5573 (2020).

[46] Sykora, P., et al. Animal recognition system based on convolutional neural network. Procedia Computer Science, 184, 346–352. (2021).

[47] Lavanya, N., & Rajesh Kumar, A. Real-Time Monitoring of Endangered Species Using AI and Renewable Edge Devices. Sustainable Computing: Informatics and Systems (2023).

[48] Sudharshan Duth P, Dhanyashree A N, Chandana P. Animal Detection and Recognition in Day Light Videos Using Deep Learning. 2024 IEEE Conference on Networking, Embedded and Wireless Systems (2024).

[49] P. Latha, Diana Inba Malar Y, Hemavarthini C S, Manjuparkavi V, Sneha S. AI-Driven Alert Systems for an Intuitive Animal Monitoring System. IEEE Conference on Advancement in Electronics & Communication Engineering (2024).

[50] Chiara De Gregorio, Raphaela Heesen, Alba Garcia-Pelegrin, et al."Quantifying facial gestures using deep learning in a New World monkey."In Scientific Reports, vol. 11, Article number: 16081, 2021.

[51] Rudresh Pillai, Neha Sharma, et al."A Deep Learning Approach for Detection and Classification of Ten Species of Monkeys."In Procedia Computer Science, vol. 173, pp. 188–195, 2020.

[52] Zhuang, P., Xing, L., Liu, Y., Guo, S., & Qiao, Y. Marine Animal Detection and Recognition with Advanced Deep Learning Models. In Working Notes of CLEF 2017 - Conference and Labs of the Evaluation Forum, CEUR-WS.org. (2017).

[53] Raja, R., & Sudha, S. Animal detection based on deep convolutional neural networks with genetic segmentation. Multimedia Tools and Applications (Springer) (2021).

[54] Rao, S., & Nikhil, B. Animal Tracking and Detection in Wildlife Videos Using YOLOv4 and Deep SORT. International Journal of Innovative Research in Computer and Communication Engineering, 9(10), 8900–8907 (2021).

[55] Sayed M. Elsayed, Rania Kora, and M. A. Elshennawy."Monkeypox Detection using Deep Learning in Informatics in Medicine Unlocked, vol. 34, 101104, 2023

[56] Kohitij Kar, Jonas Kubilius, Kailyn Schmidt, Elias B. Issa & James J. DiCarlo. "Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior." Nature Neuroscience, 22 (6), 974–983. 2019.