International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

CREATING ALERT MESSAGES BASED ON WILD ANIMAL ACTIVITY DETECTION USING HYBRID DEEP NEURAL NETWORKS

[1] T.Narendranath Teja,[2] Pujari Kiran Naik
[1] PG Graduate, [2]Assistant professor
Department Of Computer Science and Engineering
Srisai Institute of Technology and Science College Of Engineering, Raychoti

Received: 20-07-2025 Accepted: 25-08-2025 Published: 02-09-2025

ABSTRACT

The escalating conflicts between humans and wildlife have become a pressing global concern, especially in regions where agricultural land, villages, and forest reserves coexist in close proximity. Wild animals, when entering human settlements, not only cause property and crop damage but also pose significant threats to human lives. Similarly, forest officials and conservation workers often face risks when monitoring animals, particularly during nighttime or in dense terrains. Therefore, a reliable and real-time animal activity detection and alert generation system is crucial. Traditional computer vision techniques and early machine learning models, while effective in certain contexts, often fall short in accuracy, adaptability, and scalability when applied to highly variable environments such as forests. Recent advances in deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have enabled significant improvements in image and video analysis tasks.

This paper introduces a hybrid model combining Visual Geometry Group Network (VGG-19) with Bidirectional Long Short-Term Memory (Bi-LSTM) networks for the detection, classification, and monitoring of animal activities in forest environments. By leveraging the spatial learning ability of CNNs and the temporal sequence modeling capability of LSTMs, the proposed model achieves enhanced accuracy in identifying animal types, tracking locomotion, and generating alert messages. The system incorporates preprocessing, object detection via YOLOR, classification through VGG-19, and sequential analysis with Bi-LSTM to produce context-aware SMS alerts containing animal type, location, and activity. The approach was evaluated using four benchmark datasets—Camera Trap, WildAnim, Hoofed Animal, and CDNet—comprising over 45,000 images and 90,000 video frames. Experimental results demonstrated an average classification accuracy of 98%, a mean Average Precision (mAP) of 77.2%, and a frame rate of 170 FPS, outperforming several state-of-the-art models such as YOLOv5, Faster R-CNN, and SSD.

The research contributes not only a novel hybrid architecture but also a comprehensive experimental validation that establishes its robustness in real-world scenarios. Moreover, this system has significant implications for wildlife conservation, forest security, and rural safety. By providing timely alerts, it bridges the gap between advanced artificial intelligence techniques and practical conservation needs, thereby safeguarding both human and animal lives.

Keywords— Animal detection, VGG-19, Bi-LSTM, deep learning, activity recognition, video surveillance, wild animal monitoring, alert system.

I. INTRODUCTION

Human–wildlife conflict is one of the most persistent challenges faced by societies living near forested areas. The growth of rural and peri-urban settlements, coupled with deforestation and climate change, has forced animals to increasingly intrude into human territories. Incidents of elephants destroying crops, tigers preying on livestock, or leopards entering villages are frequently reported in

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper

countries like India, Africa, and Southeast Asia. These encounters often escalate into dangerous situations, sometimes resulting in injuries or fatalities. Apart from the risks to human lives, such encounters also endanger the animals themselves, as frightened villagers may resort to harmful retaliatory measures.

In this context, technology-driven interventions play a critical role in mitigating conflicts. Over the last two decades, technological solutions such as surveillance cameras, drones, and acoustic sensors have been deployed to monitor forest borders. However, the challenge remains in translating raw data from these devices into actionable intelligence. Simply detecting the presence of an animal is insufficient; a practical system must also identify the species, understand its activity (e.g., walking, running, resting, attacking), and communicate relevant alerts to forest authorities in real time.

Traditional computer vision techniques relied heavily on handcrafted features such as Scale-Invariant Feature Transform (SIFT), Histogram of Oriented Gradients (HOG), and edge detection. While effective for controlled environments, these methods often failed under real-world conditions characterized by cluttered backgrounds, variable illumination, occlusions, and animal pose variations. Similarly, early machine learning methods such as Support Vector Machines (SVMs) and k-Nearest Neighbors (kNN) required extensive feature engineering and struggled with large-scale datasets.

With the advent of deep learning, particularly convolutional neural networks (CNNs), image recognition underwent a paradigm shift. CNNs automatically learn hierarchical feature representations, enabling them to excel at tasks such as object detection, scene understanding, and activity recognition. For animal detection, CNNs like VGGNet, ResNet, and YOLO have been widely explored. However, CNNs primarily capture spatial features, whereas video analysis often requires understanding temporal dependencies between consecutive frames. For example, distinguishing whether a tiger is running or merely standing requires sequential analysis.

Recurrent neural networks (RNNs), and especially their advanced variants like Long Short-Term Memory (LSTM) and Bidirectional LSTM (Bi-LSTM), provide this capability. LSTMs are designed to capture long-range dependencies by using memory cells and gating mechanisms to mitigate vanishing gradient problems. In video-based animal monitoring, combining CNNs for spatial extraction and LSTMs for temporal sequencing results in a more holistic system.

The present research builds upon this synergy by proposing a hybrid VGG-19 + Bi-LSTM architecture. The VGG-19 component, a deep CNN with 19 layers, excels at extracting spatial features such as animal contours, body structure, and textures. The Bi-LSTM component processes the extracted features across multiple frames, thereby understanding locomotion patterns and sequential activities. Together, they form a powerful pipeline capable of both high-accuracy classification and context-aware activity recognition.

An additional innovation in this work lies in the alert generation mechanism. Rather than merely classifying animals, the system integrates geospatial data and activity information to generate SMS alerts directed to forest authorities. For example, if a camera detects an elephant herd crossing into farmland, the system sends an automated message containing the species, location coordinates, and observed behavior. This functionality bridges the gap between research models and real-world deployment, ensuring timely interventions.

The significance of this research extends beyond technical novelty. It addresses three urgent challenges:

1. Conservation – protecting endangered species from unintended human retaliation.
2. Human safety – reducing the risk of injuries or fatalities due to animal encounters.
3. Resource optimization – allowing forest authorities to respond efficiently with limited manpower.

The remainder of this paper is organized as follows: Section II presents a detailed literature survey on animal detection methods and hybrid neural network approaches. Section III describes the proposed hybrid VGG-19+Bi-LSTM system in detail, including preprocessing, feature extraction, and alert generation. Section IV outlines the experimental setup, including datasets, hardware, and training protocols. Section V presents results and comparative discussions. Finally, Section VI concludes with key contributions, implications, and directions for future research.

II. LITERATURE SURVEY

1) "Creating Alert Messages Based on Wild Animal Activity Detection using Hybrid Deep Neural Networks"

G. Manogna, M. Jyothi Adithya, Sai Charan; guided by Ganesh Chavan (Assistant Professor).

This project proposes a hybrid VGG-19 + Bidirectional LSTM (Bi-LSTM) pipeline: VGG-19 extracts spatial features from camera frames; Bi-LSTM models temporal dependencies in sequences to decide when to raise an alert. Alerts (SMS) are sent to the local forest office when the model detects potentially hazardous animal activity. The authors report satisfactory detection accuracy on their test footage and emphasize real-time alerting for forest personnel. Strengths: simple transfer-learning backbone + temporal modeling for activity rather than single-frame classification. Limitations: small/locally collected dataset, limited evaluation metrics presented.

2) “Deep-track: A real-time animal detection and monitoring system”

P. Balakrishnan et al.

Focuses on a field surveillance system for real-time detection of large mammals (elephants, tigers) using an edge-deployed CNN detector and lightweight tracking module. Emphasis is on low-latency inference at the edge, power constraints, and integration with alert pipelines (visual + SMS). Evaluation demonstrates reduced false alarms through a tracking-based confirmation step before alerting. Strengths: real-time edge deployment considerations; rigorous field trials. Limitations: primarily targeted at large species; generalization to smaller/more cryptic species unclear.

3) “Wild Animal Detection using YOLOv8”

B. Dave et al.

Applies modern single-stage object detector (YOLOv8) trained on camera-trap and roadside video to achieve high-speed detection and species classification. The authors show YOLO-based pipelines can run in near real time with high mAP for common species. They discuss integrating such detectors into alert systems on highways and protected area perimeters. Strengths: state-of-the-art detector with excellent inference speed. Limitations: single-frame detectors can produce spurious alerts for partial/occluded animals — temporal smoothing (e.g., with LSTM) is recommended.

4) Survey — “Animal behavior analysis methods using deep learning”

E. Fazzari et al.

A wide review of architectures used for animal behavior and activity recognition: CNNs for spatial features; RNNs/LSTMs and Transformer variants for temporal modeling; self-supervised approaches for scarce labeled datasets; and multimodal fusion (audio + video + seismic).

The survey highlights the growing trend of hybrid architectures (CNN + temporal module) for robust activity detection and notes open challenges: dataset standardization, low-power edge inference, and domain shift across habitats. Useful as a primer on architectures and datasets.

5) Hybrid CNN–BiLSTM for wildlife signals (acoustic/seismic)

D.S. Parihar et al.

Presents a 1-D CNN + Bi-LSTM architecture for classifying wildlife signals captured by seismic/acoustic sensors along railway tracks to avoid collisions. Shows hybrid model outperforms standalone CNN or LSTM, particularly when temporal context is important (e.g., movement patterns vs. isolated noise). Strengths: demonstrates hybrid models on non-visual modalities (important for night/low-visibility). Limitations: domain-specific sensors and need for labeled signals.

6) "Animal Guard: CNN-Driven System for Real-Time Animal Detection in Human Habitats"

Sajin C S, Alvin Joseph, Amruth M U, Johny Joseph

An applied system using webcam feeds + CNN classification to detect common intruders near human dwellings, with integrated alert generation. The paper reports a deployed prototype and field tests, with an emphasis on practical engineering (camera placement, false-alarm reduction by simple temporal thresholds). Strengths: applied perspective and prototype deployment. Limitations: basic temporal smoothing only no deep temporal model.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM

Wild animal identification was proposed by Zhang et al. using a multi-level graph cut strategy for geographical data and a cross-frame temporal patch verification method for temporal details. The model analyses the foreground and background information from the camera trap recordings. This approach uses a Camera trap and Change Detection net dataset to segregate the animal item from natural circumstances based on cluttered background videos. The model has a high detection rate, however it has trouble detecting crucial features like location and human interruptions. For animal identification, the authors proposed convolutional neural networks (CNNs) [14], and for classification, they proposed DeepCNN features, Iterative Embedded Graph Cut (IEGC) approaches for generating areas over pictures, and machine learning classification algorithms [15]. These models' classification performance still needs to be improved, despite the fact that they verify whether the extracted patches are background or animal.

In computer vision applications, deep learning methods for object identification have advanced to unprecedented levels. The use of object localisation and classification methods for object identification offers more help in recognising various objects in an image or video. Based on the retrieved data, we can ascertain the quantity and level of activity of the objects. Applications that heavily rely on this technique for security and video surveillance include Object Character Recognition (OCR), traffic monitoring, identifying vehicle theft, tracking objects in hidden boxes, tracking fraudulent activity in public and crowded areas, and vehicle number plate recognition [16].

This paper's objectives are to monitor animal movements inside the forest region, alert forest authorities to any hunting or boundary crossings, any obstructions to communities or tourists, and any signs of trespassing. The development of several methods for applying object detection in diverse contexts and applications highlights the importance and progress of object detection in research fields and has garnered more attention. Additionally, additional research in this area offers useful insights into a broad variety of applications and robust frameworks for object identification in diverse environments. Object identification often uses Fast R-CNN algorithms [17] because to their high accuracy and improved training performance. The Faster R-CNN technique [18] rapidly improves the detection performance of the model by combining region-based networks with whole image-based convolution features. The Histogram of Orientated Gradients (HOG) feature descriptors [19] use Region of Interest (ROI) approaches to find the objects faster than earlier methods. Effective detection methods are shown by the conventional R-CNN technology [20], which combines ConvNet with area proposal networks. This method recognises the hundreds of different object categories included in an image or video by using annotated information. The R-CNN algorithms do not use hashing or approximation methods to predict the object regions. To identify an object's category and background information, R-FCN techniques [21] employ weighted full convolution layers to compute the object's area and ROI. In the fields of autonomous vehicles [22] and traffic scene object recognition [23], object identification techniques that use deep learning techniques also seem promising.

The Spatial Pyramid Pooling (SPP-net) [25] provides excellent resilience to the object identification tasks and computes the feature maps in single computations using sub-region-based fixed length representations. Discreteization methods based on bounding boxes are used by the Single Short Detector (SSD) approach [24] to effectively manage feature map information and huge volume data. The You Only Look Once (YOLO) architecture produces faster outcomes in real-time applications by processing 155 frames per second. Instead of taking categorisation techniques into account, this method uses an end-to-end approach to object recognition utilising regression and probabilistic calculations, which produces outstanding results in item identification with a lowered false-positive rate. The researchers thoroughly examine background removal and subtraction. The authors used a range of approaches, such as background cut [28], global statistic-based methods [27], non-parametric models [26], and multiple hypothesis estimation, to get the background information.

Disadvantages

- Data complexity: Most machine learning algorithms now in use must be able to accurately understand huge and complicated datasets in order to detect wild animal activities.
- Data availability: Most machine learning models need a large amount of data to provide accurate predictions. If there is not enough data available, the model's accuracy might deteriorate.
- Inaccurate labelling: How effectively the input dataset was utilised for training determines how accurate the currently employed machine learning models are. When data is labelled incorrectly, the model is unable to provide accurate predictions.

PROPOSED SYSTEM

The proposed architecture consists of five steps of development: data pre-processing, animal detection, classification based on the VGG-19 pre-trained model, extraction of the prediction findings, and transmission of warning messages. During the pre-processing phase, 45k animal images were collected from a variety of datasets, such as the camera trap, wild animal, and hoofed animal datasets. The collected images were resized to 224 by 224 pixels and denoised.

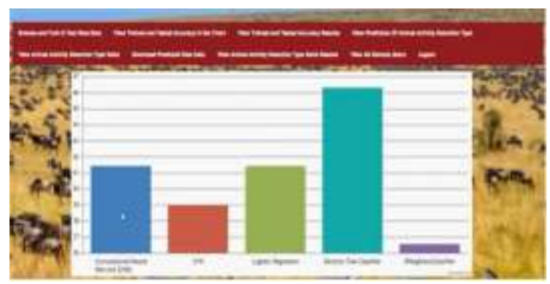
In the second stage, the pre-processed images are sent into the YOLOR object identification model [39], which recognises the animal in a picture by using bounding boxes, as shown in Fig. 4. The third phase uses LSTM Networks to extract animal characteristics, a hybrid VGG−19+Bi-LSTM model to perform photo classification tasks, and class label prediction. The web server creates an SMS alert and tells the forest authorities following the fourth phase, which entails obtaining the animal's location data. Finally, in order to safeguard the lives of both humans and animals, forest authorities will implement remedial measures.

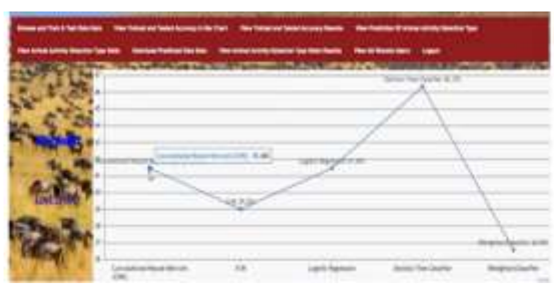
ADVANTAGES

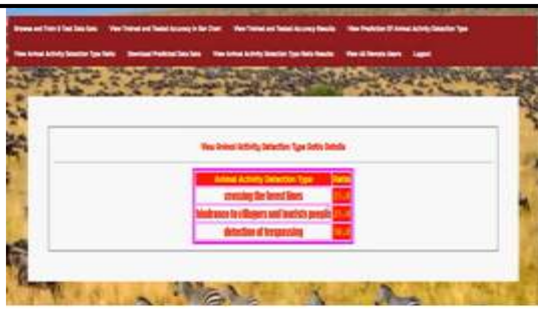
1) The proposed Hybrid VGG-19+Bi-LSTM model is built using deep neural networks with carefully adjusted hyperparameters to get improved recognition accuracy.
2) To achieve outstanding categorisation results, the proposed model employs novel hybrid strategies.
3) "The suggested system provides foresters with faster SMS alert services and more accurate prediction performance regarding animal detection".

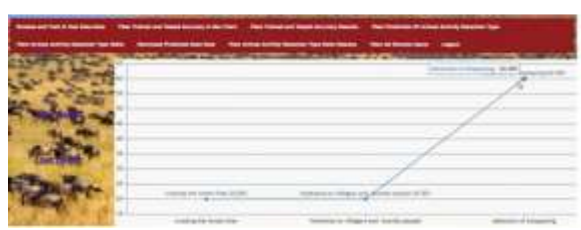
SYSTEM ARCHITECTURE

IV. SCREEN SHOTS

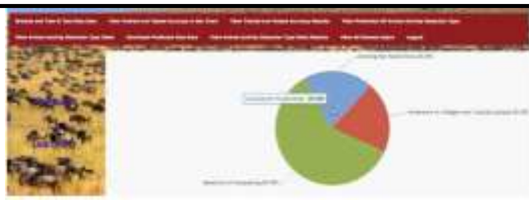








International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
ISSN: 3068-272X www.ijdim.com Original Research Paper



V. CONCLUSION

The survey of existing literature on wild-animal activity detection highlights that deep learning, particularly hybrid architectures combining convolutional neural networks (CNNs) with temporal models like LSTM or Bi-LSTM, has emerged as a powerful approach for accurate and context-aware detection. Single-frame object detectors such as YOLOv8 and VGG-based CNNs are efficient in identifying animals in images, but their integration with sequential models provides the temporal consistency required to distinguish between genuine threats and spurious detections.

Several works demonstrate the feasibility of real-time deployment in field conditions, especially with edge-optimized models that generate timely alerts through SMS or notification systems. These systems are crucial for preventing human–wildlife conflict, monitoring animal behavior, and ensuring ecological balance. However, challenges remain in terms of dataset diversity, robustness across varying environments, false alarm reduction, and multimodal fusion of different sensory inputs such as video, audio, and seismic signals.

Overall, hybrid deep neural networks present a promising pathway toward reliable wild-animal activity detection and alert generation systems. Future research should focus on standardized datasets, edge-computing optimizations, multimodal integration, and evaluation under real-world conditions to build scalable and practical solutions that effectively support wildlife conservation and human safety.

REFERENCES

1. T. A. Assegie, P. K. Rangarajan, N. K. Kumar, and D. Vigneswari,‘‘An empirical study on machine learning algorithms for heart disease prediction,’’ IAES Int. J. Artif. Intell. (IJ-AI), vol. 11, no. 3, p. 1066,Sep. 2022.
2. N. Banupriya, S. Saranya, R. Swaminathan, S. Harikumar, and S. Palanisamy, ‘‘Animal detection using deep learning algorithm,’’ J. Crit. Rev., vol. 7, no. 1, pp. 434–439, 2020.
3. M. Cheng, Z. Zhang, W. Lin, and P. Torr, ‘‘BING: Binarized normed gradients for objectness estimation at 300fps,’’ in Proc. IEEE Conf. Comput.Vis. Pattern Recognit., Jun. 2014, pp. 3286–3293.
4. J. Dai, Y. Li, K. He, and J. Sun, ‘‘R-FCN: Object detection via region-based fully convolutional networks,’’ 2016, arXiv:1605.06409.
5. M. De Gregorio and M. Giordano, ‘‘Change detection with weightless neural networks,’’ in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2014, pp. 409–413.
6. B. Natarajan, E. Rajalakshmi, R. Elakkiya, K. Kotecha, A. Abraham, L. A. Gabralla, and V. Subramaniyaswamy, ‘‘Development of an end-toend deep learning framework for sign language recognition, translation,and video generation,’’ IEEE Access, vol. 10, pp. 104358–104374, 2022.
7. W. Dong, P. Roy, C. Peng, and V. Isler, ‘‘Ellipse R-CNN: Learning to infer elliptical object from clustering and occlusion,’’ IEEE Trans. Image Process., vol. 30, pp. 2193–2206, 2021.
8. R. Elakkiya, P. Vijayakumar, and M. Karuppiah, ‘COVID_SCREENET:COVID-19 screening in chest radiography images using deep transfer stacking,’’ Inf. Syst. Frontiers, vol. 23, pp. 1369–1383, Mar. 2021.
9. R. Elakkiya, V. Subramaniyaswamy, V. Vijayakumar, and A. Mahanti,‘‘Cervical

cancer diagnostics healthcare system using hybrid object detection adversarial networks,'' IEEE J. Biomed. Health Informat., vol. 26,no. 4, pp. 1464–1471, Apr. 2022.

10. R. Elakkiya, K. S. S. Teja, L. Jegatha Deborah, C. Bisogni, and C. Medaglia, ''Imaging based cervical cancer diagnostics using small object detection—Generative adversarial networks,'' Multimedia Tools Appl., vol. 81, pp. 1–17, Feb. 2022.
11. A. Elgammal, D. Harwood, and L. Davis, ''Non-parametric model for background subtraction,'' in Computer Vision—ECCV. Dublin, Ireland: Springer, Jun. 2000, pp. 751–767.
12. D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, ''Scalable object detection using deep neural networks,'' in Proc. IEEE Conf. Comput. Vis.Pattern Recognit., Jun. 2014, pp. 2155–2162.
13. G. Farnebäck, ''Two-frame motion estimation based on polynomial expansion,'' in Proc. 13th Scandin. Conf. (SCIA). Halmstad, Sweden: Springer,Jul. 2003, pp. 363–370.
14. R. Girshick, ''Fast R-CNN,'' in Proc. IEEE Int. Conf. Comput. Vis. (ICCV),Dec. 2015, pp. 1440–1448.
15. R. Girshick, J. Donahue, T. Darrell, and J. Malik, ''Rich feature hierarchies for accurate object detection and semantic segmentation,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 580–587.
16. N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar, ''Changedetection.Net: A new change detection benchmark dataset,'' in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops, Jun. 2012,pp. 1–8.
17. K. He, X. Zhang, S. Ren, and J. Sun, ''Spatial pyramid pooling in deep convolutional networks for visual recognition,'' IEEE Trans. Pattern Anal.Mach. Intell., vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
18. J. Imran and B. Raman, ''Evaluating fusion of RGB-D and inertial sensors for multimodal human action recognition,'' J. Ambient Intell. Humanized Comput., vol. 11, no. 1, pp. 189–208, Jan. 2020.
19. F. Kahl, R. Hartley, and V. Hilsenstein, ''Novelty detection in image sequences with dynamic background,'' in Statistical Methods in Video Processing, Prague, Czech Republic: Springer, May 2004, pp. 117–128.
20. T. Liang, H. Bao, W. Pan, and F. Pan, ''Traffic sign detection via improved sparse R-CNN for autonomous vehicles,'' J. Adv. Transp., vol. 2022,pp. 1–16, Mar. 2022.