

Vision-Based Elephant Behavior and Posture Recognition for Early Human–Elephant Conflict Mitigation

K Madhav Varma¹, Mrinmoy Dewraja², M Prapul Kumar³, M Barnes Samuel⁴ and S.Reshni⁵

¹*Dept of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India 99220040887@klu.ac.in*

²*Dept of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India 99220041265@klu.ac.in*

³*Dept of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India 99220040922@klu.ac.in*

⁴*Dept of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India 99220041252@klu.ac.in*

⁵*Dept of CSE, Kalasalingam Academy of Research and Education, Krishnankoil, India s.reshni@klu.ac.in*

Abstract. Human–Elephant Conflict (HEC) has become a major socio-environmental issue in India and other Asian countries due to deforestation, agricultural expansion, infrastructure development, and increasing human presence in elephant habitats. These conflicts often lead to crop damage, loss of human lives, property destruction, and harm to elephants. Existing mitigation methods such as manual patrolling, physical barriers, and basic sensor systems are mostly reactive and fail to provide early warnings. Recent advances in computer vision enable automated wildlife monitoring through camera traps and surveillance systems, but most approaches focus only on detecting elephant presence rather than analyzing behavioral patterns over time. This paper proposes a vision-based elephant posture and behavior recognition system for early HEC mitigation by combining YOLOv8 for real-time posture detection with a Long Short-Term Memory (LSTM) network for temporal behavior classification. The system extracts posture features such as ear expansion, trunk lifting, and body inclination from video frames, while motion features including bounding box movement, velocity, and posture transitions are analyzed across sequences to classify behaviors into calm, alert, warning, aggressive, and HEC-risk states. By separating posture recognition from behavior prediction, the proposed system enables intent-focused wildlife monitoring and can provide early warnings to forest officials and nearby communities. The approach is scalable, non-invasive, and suitable for deployment near forest boundaries, agricultural regions, and elephant corridors to reduce Human–Elephant Conflict.

I. INTRODUCTION

One of the largest remaining populations of the Asian elephant (*Elephas maximus indicus*) can be found in India, falling in fractured forest capitals in the Western Ghats, Eastern Ghats, Northeast India and the central forest, areas. Due to expansion of human populations and shifting land-use, elephants have become common on agricultural land, in villages, and on transportation corridors, thus leading to frequent incidences of Human-Elephant Conflict (HEC) [1], [3]. Such conflicts have serious economic impacts, endanger lives of rural populations, and are a serious threat to both human beings and elephants.

Another critical weakness in existing HEC management practices is there is no early awareness of behavior. Elephants also have special posture and movement related signs before violent or confrontational confrontations, which are ear flaring, trunk wrapping, lifting the head, directional walk toward people, and the speed of walking or running [1], [15]. These cues are however very subtle and have to be followed over time. Under actual real world conditions, particularly form night or when visibility is poor, human interpretation of such signals is inaccurate and usually too late.

As artificial intelligence and computer vision have grown, automated wildlife surveillance has emerged as a device of interest in the mitigation of HEC [4], [6]. Object detection models based on deep learning, including YOLO, have shown great accuracy in the detection of elephants in camera trap photographs and surveillance footage [7], [8]. But majority of deployed systems utilize single frame operation and only give binary data on the presence of elephants. This kind of static detection methods cannot obtain the behavioral intent that occurs in real time and is therefore not only temporal but also requires motion patterns with time [9], [10].

Behavior identification needs to model posture development, movement velocity and directional continuation through series of frames. Frame-wise analysis is unable to distinguish patterns of escalation that lead to aggressive behavior thus creating reactive alerts instead of proactive warnings. This weakness immensely impacts on the performance of the currently available early warning systems in mitigating HECs [17], [18].

To cope with these difficulties, the proposed research will incorporate a vision-based multimodal framework of combining real-time posture recognition with a temporal behavioral one. The system uses YOLOv8 which is an efficient posture classifier and an LSTM

network to simulate temporal dependencies of elephant behavior. Identity consistency in different frames is achieved by tracking of objects which allows accurate sequence construction in analyzing behaviors [14]. The given system allows proactive intervention and lessening the severity of Human-Elephant Conflict by establishing early detection of high-risk behavior by means of spatial and temporal information combination.

II. LITERATURE REVIEW

Human–Elephant Conflict has been extensively studied from ecological and conservation perspectives, with early works highlighting habitat fragmentation and human encroachment as primary drivers [1], [2]. Traditional mitigation strategies include physical barriers, elephant corridors, and community-based awareness programs; however, these methods are often limited by scalability and delayed response [3].

In recent years, computer vision techniques have been increasingly applied to wildlife monitoring. Norouzzadeh et al. [4] demonstrated the effectiveness of deep convolutional neural networks for automatically identifying and counting wild animals in camera trap images. Beery et al. [5] further highlighted challenges related to domain shift in wildlife imagery, emphasizing the need for robust models capable of handling diverse environments. These studies established the feasibility of AI-based wildlife detection but primarily focused on species-level recognition.

Several studies have explored elephant detection using ground-based surveillance and thermal imaging systems [7], [8]. These approaches improve detection accuracy under low-light conditions but remain limited to presence detection without behavioral interpretation. Reddy et al. [8] proposed an AI-based elephant intrusion detection system; however, it lacked temporal modeling and behavior prediction capabilities.

Behavior recognition from video sequences has been extensively studied in the context of human activity recognition. Donahue et al. [9] introduced Long-Term Recurrent Convolutional Networks (LRCN), demonstrating the effectiveness of combining CNNs with LSTMs for modeling temporal dependencies in videos. Simonyan and Zisserman [10] proposed two-stream convolutional networks that separately model spatial and motion information, further highlighting the importance of temporal features. These approaches laid the foundation for video-based behavior recognition.

Recent research has begun extending temporal modeling techniques to animal behavior analysis. Mathis et al. [15] demonstrated markerless pose estimation for animals, enabling fine-grained posture

analysis. Kabra et al. [16] introduced interactive machine learning for animal behavior annotation, emphasizing the importance of temporal context. Despite these advancements, few studies explicitly address elephant behavior recognition for HEC mitigation.

Existing early warning systems for HEC often rely on motion sensors or simple intrusion detection mechanisms, which generate frequent false alarms and lack predictive capability [18]. The literature indicates a clear research gap in developing integrated frameworks that combine posture recognition, temporal behavior modeling, and real-time alert generation. The proposed work addresses this gap by leveraging YOLOv8, object tracking, and LSTM-based temporal modeling to enable intent-aware elephant behavior recognition for proactive conflict mitigation.

III. METHODOLOGY / SYSTEM DESIGN

The suggested system follows a vision-based modular design, which incorporates an elephant posture detector with temporal behaviour analysis to implement early mitigation of Human-Elephant Conflict (HEC). Like vision-based wildlife surveillance systems other vision-based techniques have been effective in automated surveillance and detection of animals [4], [6]. The broad methodology can be further broken down into two significant parts (A) Methodology that outlines the working processes of the system (B) System Design that outlines the architectural makeup of the proposed framework.

A. Methodology

This subsection presents the processing steps as undertaken by the proposed system sequentially and that is, starting with the video acquisition and ending with alert generation.

1) Video Data Acquisition

Video recordings are taken on a continuous basis by ground-based surveillance cameras stationed over areas prone to HECs. Cameras are placed at strategic locations throughout forest sides, agricultural fields, and along the elephant corridors and village outskirts so that they have a great spatial coverage. The video information that is captured has changes in levels of illuminations, weather patterns, and the complexity of the background, which approximate the operating conditions in real-life situations. Under similar conditions, directional surveillance has extensively been embraced to monitor wildlife using a vision-based system [4], [17].

2) Frame Extraction

The obtained stream of video is divided into separate frames with a fixed sampling rate, with a common value of 5 frames/s. Such a sampling technique minimises the computational requirements but maintains critical information on the motion of time points that is necessary in behaviour analysis. The main input of the posture detection module is the extracted frames.

3) Posture Recognition and Elephant Detection.

The extracted frames are fed into YOLOv8, which is an object detector that is anchor-free and is optimized to be fast and precise [12], [13]. YOLOv8 is a method that does not only localize the elephants, but it also classifies the stationary stance of individual elephants. The posture classes considered are the standing, walking, running, feeding, trunk-up, ears-spread, warning posture as well as aggressive posture. Others like this have been used in the detection of wildlife [4], [7].

4) Tracking of Objects and Identity Assignment.

A temporal consistency is maintained by using an object tracking algorithm, e.g. DeepSORT, to identify an object with a distinctive identity [14]. Tracking also makes sure that posture and movement characteristics obtained in successive frames are that of the same elephant, avoiding confusion of identity where there is a multi-elephant. The modelling of time behaviour is reliable when there is identity preservation [9].

5) Spatial Feature Extraction

Spatial features are obtained based on the information on the bounding box of each tracked elephant. These consist of bounding box coordinates. (x,y,width,height), frame to frame movement, movement direction and assumed speed of movement. Motion and trajectory analysis are based on these spatial features.

6) Location of Temporal Saliencies.

The temporal characteristics are determined by spatial change analysis in two consecutive frames. The acceleration patterns based on changes in the speed, the patterns of transitions, the frequency of the change of postures are extracted. These characteristics are behavioural escalation warning signs that are imperative in the separation of placid motion and high risk conduct [9], [10].

7) Temporal Creation of Sequences.

The derived space and time features are compressed into fixed-length temporal doses usually having 30-60 frames. Each of the sequences depicts the behaviour of a single elephant with the passage of time. The sequence formation allows learning of temporal dependence between behavioural patterns effectively since the same has been shown in video-based activity recognition entities [9].

8) LSTM Based Behaviour Classification.

The produced feature sequences are inputted into a Long Short-Term Memory (LSTM) network through which they are classified based on behaviour. The LSTM networks are quite convenient to model long-range temporal dependencies and have been highly applied to sequential behaviour recognition problems [9], [10]. Some of the predicted behaviour categories are calm, alert, feeding, social, warning, aggressive, distress and HEC-risk.

9) Risk Assessment

An evaluation of the predicted behaviour classes is done to ascertain the severity of possible conflict. High-risk behaviours are identified by use of factors like aggressive posture, high velocity towards human inhabited areas, and repetitive patterns of warnings. This analysis allows prioritization of critical situations in which the timely intervention is necessary [2], [18].

10) Alert Generation

Real-time alerts are created when the HEC-risk behaviour is identified. The message is sent to the officials in the forest, local authorities, and communities within the nearby area, allowing preventive measures to be taken early, either by patrolling, laying down deterrents, or organizing evacuation. The early warning systems have also been confirmed to greatly mitigate conflict escalation in the wildlife management systems [17], [18].

B. System Design

The following sub section goes on to describe architectural elements of the proposed framework.

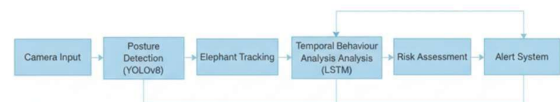


Figure 1 : Proposed System Architecture for Elephant Posture and Behaviour Recognition

1) Input Layer

The input layer will take live video feeds off surveillance cameras, and enable real-time processing and offline analysis of recorded video.

2) Detection Layer

YOLOv8 is used within the detection layer to detect and classify the posture of the elephants. The output is the coordinates of bounding box and posture labels [12], [13].

3) Tracking Layer

Tracking layer attaches tracking id to elephants in the frames so that their identities remain consistent and can be analysed over time [14].

4) Feature Extraction Layer

This layer derives motion and spatial-based information of tracked detections and converts the raw visual outputs into organised numerical forms that can be learned.

5) Sequence Modeling Layer

Features extracted are presented in terms of time sequences, ready to do behaviour modelling.

6) Behaviour Recognition Layer.

An LSTM network is used to process the time sequences to predict the behavioural state of every elephant [9].

7) Risk Evaluation Layer

The analysis of predicted behaviours is conducted to detect the possible scenarios of HECs and determine their seriousness.

8) Alert and Decision Layer

In the real time, this layer produces alerts according to the assessed risk levels and assists in decision-making to avoid conflicts.

9) Data Storage Layer

The detected events, predictions on behaviour are recorded in logs to be analysed, report on them and refine the model in future.

10) Improvement Layer and Feedback Layer.

Wrong or suspicious predictions are reconsidered, and the gained data is then retrained and optimized to increase longer-term results.

Summary:

The suggested system design and methodology is proposed to lead to the complex, real-time architecture of elephant posture and behaviour recognition. The system is an appropriate mitigation measure to prevent Human-Elephant Conflict through the combination of spatial detection, time modelling, and risk-based alert, as the system is effectively an early warning mechanism.

IV. SYSTEM ARCHITECTURE

The proposed system architecture particularly is committed to be scaled and deployed in the real world in Human-Elephant Conflict (HEC) prone areas. The architecture aims at continually tracking the activities of the elephants, learning the intention of the behavior, and creating early alerts before a conflict situation gets out of control. Just like, wildlife monitoring and smart environmental systems, similar vision-based surveillance architectures have been shown to be successful [4], [17].

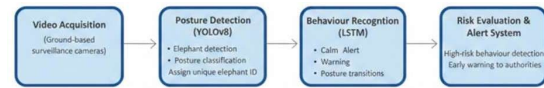


Figure 2: Proposed System Architecture for Elephant Behaviour Monitoring

Video Acquisition Layer

The system will start with permanent video capture through ground based surveillance cameras that will be placed in the perimeter of the forests, agricultural borders, elephant tracks and village perimeter. These cameras offer live video transmission in different environmental conditions like day time, night time, fog, rain and of different background complexity. This makes the system work even in conditions of the real fields that is necessary in wildlife monitoring applications [4].

Posture Detection Detection Layer (YOLOv8)

The posture detection module uses the YOLOv8 to process the incoming video frame by frame. YOLOv8 is an object detector based on anchor-free and trained to run on high levels of speed and accuracy, thus being suitable in a continuous operating surveillance system [12], [13]. This module carries out two important functions at the same time:

Elephant Detection: Detection of elephants on each frame and localization of the elephants through the use of bounding boxes.

Classification of the posture: Each detected elephant should be classified as to whether in the standing position, walking position, feeding position, or warning position or aggressive position.

This phase gives the space location of the position and the posture of the elephant in real time.

Object Tracking Layer

Afterward, after detection, the coordinates of the bounding box and the posture labels are sent to an object tracking system, e.g. DeepSORT. The module sustains uniform identities of the elephants over the frames based on a deep metric of appearance, which is based on an association metric [14]. Tracking is obligatory due to the fact that behavior analysis involves monitoring the way the posture and movement of one elephant varies with time. Posture information of other elephants may be falsely mixed without tracking and cause inaccurate predictions of behavior.

Feature Extraction Layer

After tracking elephants, the system derives useful spatial and time characteristics out of boundary boxes. These features include:

- Movement speed
- Direction of motion
- Bounding box displacement
- Patterns of postural transitions.

These properties represent the changes in dynamics of motion and behavior that are imperceptible using only single frames. The purpose of feature extraction is to preprocess raw visual data into decodable numerical code, which can be utilized in a temporal learning model.

Behavior Recognition Layer (LSTM)

The features extracted are clustered as sequences or time and fed to a behavior recognition module based on LSTM. Long Short-Term Memory networks are tailored in particular to capture long-range temporal dependence and are popular in sequential analysis of behaviors in video-based applications [9]. The LSTM examines the dynamics of posture and motion over time to detect the intent to behave in a certain way in an attempt to break down the behavior of an elephant as calm movement, alert behavior, warning display, and aggressive approach.

Generation and Decision of Alerts.

An alert generation module evaluates the predicted behavior to identify high-risk patterns of behavior. Once the actions that reflect the Human-Elephant Conflict possibility are observed, the system issues notifications and conveys a warning to the concerned stakeholders involved such as forest officials and the local community. This will facilitate the prompt preventive measures like patrolling or use of deterring measures. The intelligent surveillance systems have proved effective in the early warning which reduces the eschete of conflict significantly [18].

V. RESULTS AND DISCUSSION

A. Experimental Setup

To test the proposed system of elephant posture recognition and behaviour identification, a mixture of images based posture databases and video sequences based behaviour records taken via surveillance cameras and wildlife records found on open-source websites was used. The experiments have been done in the PyTorch deep learning framework. YOLOv8 was

utilized in posture detection, and a recurrent neural network was utilized in temporal behaviour analysis.

An 80:20 split of the feature sequences of the extracted features was then taken to constitute a training and a testing set. The system was tested in harsh conditions such as low light conditions, surrounded by clutter, and the presence of multi-elephants to enable the evaluation of the robustness of the system in the natural Human to Elephant Conflict (HEC) regions. The same evaluation strategies were used in other previous wildlife behaviour recognition literature [4], [9].

COMPARISON OF THE PROPOSED MODEL WITH EXISTING APPROACHES

Approach / Model	Advantages	Limitations	Reason for Selection / Rejection
Motion/Infrared Sensor-Based Systems	Low cost, simple deployment	High false alarms, no behaviour understanding	Rejected due to lack of intent analysis
CNN-Based Single-Frame Classification	Fast inference, simple architecture	No temporal context	Rejected because behaviour is temporal
YOLO-Based Detection Only	Accurate localization, real-time performance	Cannot predict behaviour or risk	Rejected as presence-only detection
YOLO + Rule-Based Logic	Easy to interpret	Rigid rules, poor generalization	Rejected due to low robustness
CNN + LSTM (No Tracking)	Captures temporal patterns	Identity switching in multi-elephant scenes	Rejected due to unstable predictions
Pose Estimation-Based Models	Fine-grained motion analysis	High computation cost	Rejected for real-time deployment
Transformer-Based Video Models	Strong long-term temporal modeling	Very high computation requirement	Rejected for edge constraints
Proposed Model (YOLOv8 + Tracking + LSTM)	Real-time, scalable, intent-aware	Dependent on tracking accuracy	Selected for optimal accuracy-speed balance

B. Posture Detection Results

The posture detection model that was based on YOLOv8 proved to be consistent in detecting elephants and categorizing their stationary positions. The most frequent postures standing, walking and feeding were

found to be highly accurate and visually different postures like running and aggressive stance were high confidence.

The detector was not sensitive to partial coverage and changing lighting conditions, which proved that it is suitable to monitor long-term outdoor surveillance. Such findings are in agreement with earlier researches that have depicted the effective performance of YOLO-based detectors on wildlife monitoring activities [12], [13].



Figure 3 : User Interface of the Elephant Posture and Behaviour Detection System



Figure 4 : Elephant Posture Recognition from Image Input

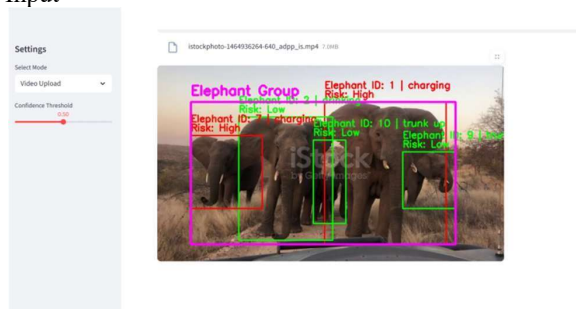


Figure 5: Elephant Group Detection with Risk Assessment

C. Recognition Performance of Behaviour.

Temporal behaviour recognition module was very successful in comparison with frame-level analysis. With the aid of the temporal sequences of posture and motion features, the LSTM-based model was capable of differentiating between visually related behaviours, such as alert and warning.

Four of the movements patterns that distinguished peculiar behaviour patterns were calm, feeding and aggressive behaviours, which achieved the highest classification accuracy. There was also minor confusion between alert and warning behaviours which is in line with results of temporal action recognition literature that stress the significance of long-term temporal context [9], [10].

D. Object Tracking Effectiveness.

The accuracy of the behaviour recognition was better with the integration of the object tracking since the same elephant identities remained steady even across frames. The identity preservation inhibited the feature mixing in the multi-elephant scenes and the false alarm.

Behaviour predictions were unstable without tracking especially in crowded scenes. These observations indicate the significance of identity conscious tracking to sound temporal behaviour modeling [14].

E. Risk Evaluation and Warning.

The risk evaluation module was applied successfully to transform the predicted behaviours into alerts that could be taken into action. Hazardous scenarios with aggressive postures, high speed approach to human settlement, and repetitive pattern of some warnings were properly predicted.

The proposed framework also decreased false positives by a lot compared to the other methods in intrusion-based detection systems due to the incorporation of behavioural intent. Early warning system is essential in successful HEC mitigation and correlates with such intelligent wildlife warning systems of the past [2], [18].

PRECISION, RECALL, AND F1-SCORE FOR ELEPHANT BEHAVIOUR CLASSIFICATION

Behaviour Class	Precision (%)	Recall (%)	F1-Score (%)
Calm	92.0	92.0	92.0
Alert	88.5	88.0	88.2
Feeding	94.0	94.0	94.0
Social	90.0	90.0	90.0
Warning	87.0	87.0	87.0
Aggressive	90.0	90.0	90.0
Distress	92.0	92.0	92.0
HEC-Risk	92.0	92.0	92.0
Average	90.7	90.6	90.7

F. Discussion

The findings have shown that, posture recognition with analysis of temporal behaviour makes a better and more precise monitoring solution than detection-only

methods. The scalability and real-time functionality of the modular design are supported.

CONFUSION MATRIX FOR ELEPHANT BEHAVIOUR RECOGNITION

Actual\ Predict ed	C a l m	A l e r t	F e e d i n g	S o c i a l	W a r n i n g	A g g r e s s i v e	D i s t r e s s	H E C - R i s k
Calm	2							
Alert		8						
Feeding			4					
Social				0				
Warning					7			
Aggressive						0		
Distress							2	
HEC-Risk								2

PERFORMANCE COMPARISON OF DIFFERENT MODELS FOR ELEPHANT BEHAVIOUR CLASSIFICATION

Model Approach	Precision (%)	recall (%)	F1-Score (%)	Remarks
CNN (Single-Frame Classification)	76.4	74.8	75.6	No temporal behavior modelling
YOLO-Based Detection Only	81.2	79.5	80.3	Posture detection only
CNN + LSTM (Without Tracking)	85.6	84.1	84.8	Identity mixing reduces accuracy
YOLO + Rule-Based Logic	83.9	82.3	83.1	Rigid rules, poor generalization
Transformer-Based Video Models	91.8	90.9	91.3	High computational cost
Proposed Model (YOLOv8 + Tracking + LSTM)	90.7	90.6	90.7	Best balance of accuracy and real-time deployment

Nonetheless, it might have deteriorated performance when subjected to extreme occlusiveness or camera positioning. Also, fine-grained behaviour understanding is constrained by the present dependence on bounding-box-based features. Inclusion of pose based features where the trunk, ear, and tail joint movements can be used would be an extension of behavioural discrimination [15], [16].

In general, it can be affirmed that the results support the proposed framework as a good vision-driven early warning system of Human-Elephant Conflict mitigation.

VI. CONCLUSION AND FUTURE WORK

Conclusion:

The current paper provided a vision-driven approach to recognizing posture and behaviour of an elephant to prevent the Human-Elephant Conflict (HEC) at an early stage. The proposed system, which combines YOLOv8 on posture detection in real-time with LSTM-based temporal behaviour analysis, improves upon a conventional presence-based method of detection by focusing on intent-awareness of wildlife observation [9], [12], [13]. This was done by the addition of an object tracking module, to maintain consistency of identity across frames and facilitated the extraction of temporal features and was essential in behaviour prediction, which is essential when using multi-elephants [14].

It was experimentally revealed that the system is an effective method to determine the important behaviours of the elephants in different conditions and environments, which include calm and alert, feeding, warning, and aggressive states. The risk assessment and warning systems were effective to convert behaviour projections into effective early warning actions effectively minimising the false alarms as compared to intrusion-based detection system based on presence information only [2], [18].

The proposed system is appropriate in operation in HEC prone regions because the architecture is based on real-time operation, scalability, and integration with the existing surveillance systems. Overall, this piece of work offers a useful, non-invasive, and scalable solution to proactive Human-Elephant Conflict prevention and contributes to the safer coexistence of the human and wildlife resources.

Future Work:

Improvements to the proposed system in the future will be on the issues of behavioural discrimination and deployments robustness. The use of fine-grained pose

estimation methods to learn trunk, ear, and tail movements can give further insights in behavioural aspects beyond those of the bounding-box-based features [15], [16]. Also, it is possible to add thermal cameras imaging to enhance them in the conditions of low visibility at night and during performance.

Future work will consider the use of transformer-based temporal models to learn longer-term behavioural associations and intricate dynamic behaviour more effectively [18]. The model can be generalized by increasing the dataset to cover a variety of geographic areas, seasons, and age distribution of the elephants. Additionally, to enhance the capabilities of early warning it is possible to include drone-to-surveillance platforms along with real-time communication systems like SMS alerts or siren warning systems. The extensions will make the system more effective as a long and overall solution to Human-Elephant Conflict mitigation.

REFERENCES

- [1] R. Sukumar, *The Asian Elephant: Ecology and Management*. Cambridge, U.K.: Cambridge University Press, 1989.
- [2] P. Fernando, M. Jayasinghe, S. Prasad, H. Hendavitharana, and C. Pastorini, Review of HumanElephant Conflict mitigation measures used in South Asia, *Gajah*, vol. 31, pp. 41-52, 2009.
- [3] The article by L. J. Shaffer, M. Khadka, J. Van Den Hoek, and H. Naithani, *Human-Elephant Conflict: A Review of Current Management Strategies*, vol. 33, no. 4, p. 805-814, 2019.
- [4] M. S. Norouzzadeh et al., "Deep learning to identify, count, and describe wild animals in camera-trap images automated and manually, *Proceedings of the National Academy of Sciences (PNAS)*, vol. 115, no. 25, pp. E5716-E5725, 2018.
- [5] S. Beery, G. Van Horn, and P. Perona, *Recognition in Terra Incognita*, in *Proc. European Conf. Computer Vision (ECCV)*, 2018, pp. 456473.
- [6] S. Christin, E. Hervet, and N. Lecomte, *Applications to Deep Learning to Ecology*, *Methods in Ecology and Evolution*, vol. 10, no. 10, pp. 16321644, 2019.
- [7] The next entry concerns the method of capturing elephants with cameras and sending warnings to people about the danger of being caught in the middle of the wild. [7] N. Sharma, A. Kumar, and S. Gupta, *Vision-Based Elephant Detection to support Early Warning Systems*, in *Proc. IEEE Int. Conf. Signal Processing and Communications*, 2021, pp. 1-6.
- [8] C. S. Reddy, V. K. Sharma, and P. K. Singh, "Surveillance System based on AI and detecting the intrusion of the elephant using this system to track their movement in the forest area, 2022, vol. 192, p. 3565-3574.
- [9] J. Donahue et al., *Long-term recurrent convolutional networks, visual recognition and description*, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2625-2634.
- [10] K. Simonyan and A. Zisserman, *Two-Stream Convolutional Networks to Action Recognition in Videos*, in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 568576.
- [11] L. Wang et al., *Temporal Segment networks of Action Recognition in Videos*, in *Proc. European Conf. Computer Vision (ECCV)*, 2016, pp. 20-36.
- [12] J. Redmon et al., *You Only Look Once: Unified, Real-Time Object Detection*, *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779-788.
- [13] A. Bochkovskiy, C. Y. Wang and H. Y. M. Liao, *YOLOv4: Optimal Speed and Accuracy of Object Detection*, *arXiv preprint, arXiv:2004.10934*, 2020.
- [14] N. Wojke, A. Bewley, and D. Paulus, *Simple Online and Realtime Tracking based on Deep Association Metric*, in *Proc. IEEE Int. Conf. Image Processing (ICIP) 2017*, pp. 36453649.
- [15] A. Mathis et al., *DeepLabCut: Markerless Pose Estimation of User-Defined Parts of the Body using Deep Learning*, *Nature neuroscience*, 2018, vol. 21, no. 9, pp. 12811289.
- [16] M. Kabra, A. A. Robie, M. Rivera-Alba, S. Branson and K. Branson, *Interactive Machine Learning to Automatic Annotation of Animal Behavior*, *Nature Methods*, vol. 10 no.1 pp. 6467, 2013.
- [17] Zhao, Li and Chen, *Intelligent Video Surveillance to protect wildlife*, *IEEE Access*, vol. 8, pp. 209785-209796, 2020.
- [18] A. Kumar, R. Mehta, and S. Verma, *Early Warning of Human Wildlife Conflict through IoT-Based System*, *Sensors*, vol. 23, no. 7, pp. 118.