Email: editor@ijerst.com or editor.ijerst@gmail.com

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015 Vol. 21, Issue 2, 2025

A HYBRID DEEP LEARNING FRAMEWORK FOR WILD ANIMAL DETECTION AND AUTOMATED ALERT MESSAGING

[1]Ponuganti Jaya Chandra Rao, MCA Student, Department of MCA

[2]P. Rohini Bai, M.Tech, Assistant, Professor, Department of MCA

[12]Dr KV Subba Reddy Institute of Technology, Dupadu, Kurnool

ABSTRACT

Forestry workers and rural residents are becoming more and more concerned about the problem of animal assaults. Both drones and surveillance cameras are often used to monitor the movements of wild animals. To identify the sort of animal, track its movement, and offer its position, an effective model is necessary. The safety of both people and foresters may then be guaranteed by sending alert signals. Although methods based on computer vision and machine learning are widely employed for animal identification, their cost and complexity make it challenging to get adequate results. In order to identify animals and provide warnings depending on their behaviour, this research proposes a Hybrid Visual Geometry Group (VGG)−19+ Bidirectional Long Short-Term Memory (Bi-LSTM) network. To enable prompt action, these notifications are sent by Short Message Service (SMS) to the nearby forest office. With an average classification accuracy of 98%, a mean Average Precision (mAP) of 77.2%, and a Frame Per Second (FPS) of 170, the suggested model demonstrates significant gains in model performance. Using 40,000 photos from three distinct benchmark datasets with 25 classes, the model was evaluated both subjectively and statistically. Its mean accuracy and precision were higher than 98%. This approach is a dependable way to save human life and provide precise information based on animals.

I. INTRODUCTION

Generally speaking, the constant stream of information and the crowded backdrops make it difficult for researchers to spot animal behaviour. Numerous animal groups exist, each having a unique set of features on its face, snout, body, and tail. A strong framework must be created in order to analyse large feature maps and recognise and classify such creatures in video sequences. Such advancements in real-time scenarios need significant GGPU-based processing capabilities as well as vast amounts of video data for testing and training. Furthermore, in order to provide believable findings, the integrating approaches should treat the data intelligently. The need for creating such a program to identify animal activity in forested areas is thus quite great. Even though this technology age has seen many breakthroughs, more study is still needed in this field to develop a solid model. By doing this, we can protect people from unexpected animal assaults and provide forest authorities with alarm messages that include location data so they can respond quickly. Better monitoring services are provided by these devices, which also assist in tracking animal activity and identifying instances of human hunting or wildlife disturbance. These activity clusters provide significant challenges in the field of deep learning, such as monitoring an animal item, determining its activity, and producing alarm signals. This work's research explores the developments in sophisticated neural network-based architectures and video analysis algorithms. Impressive outcomes in picture identification, classification, and creation

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015

Vol. 21, Issue 2, 2025

tasks have been achieved by recent advancements in Deep Learning approaches [1]. As a result of these advancements, we concentrate on creating a reliable model for tracking animal behaviour and notifying forest officials of any unusual activity, such as hunting or animals intruding into agricultural or residential areas. The creation of the suggested model looks at this issue from many perspectives in order to provide a more effective answer.

Techniques for object recognition are essential for comprehending the elements of pictures and the connections between them. It expressly gives the movement and activity-based data in the case of videos. The traditional approaches provide observable outcomes and use manually designed processes [2], [3] for feature extractions. This work is efficiently handled via the creation of deep learning models, which lowers the overheads seen in previous research. Traditional machine learning techniques were used in earlier research to recognise objects, however they run into problems when dealing with multimodal inputs and large datasets. The deep model integrates the pertinent features to create feature maps and efficiently manages an image's features to investigate the finely tailored inquiry on pixels. Without the need for human interaction, feature maps make it simple to understand an object's structure and anticipate its patterns, curves, edges, and contours. Deep learning models are designed to scale massive amounts of data and manage such intricate data structures. To get high accuracy outcomes, the deep neural network performance is regulated by regularisation and hyperparameter optimisation strategies. Object detection algorithms are often used in a variety of disciplines, including salient object identification [7], [8], scene understanding [6], and face detection [4], [5].

Research on the detection of animal activity is still in its early stages. For the preceding methods to provide believable results, they must be improved and adjusted. A total of 40K photos from four distinct benchmark datasets were utilised. Using appropriate picture sizes and quality criteria, the suggested model has been assessed both subjectively and quantitatively. However, it's clear that combining several deep networks in a hybrid fashion increases the model development's complexity overhead. On a mix of object identification and class prediction tasks, however, the effective application of such algorithms yields unthinkable outcomes [9]. In light of this, we suggest a new method called "hybrid VGG−19+Bi-LSTM networks" to identify animal activity and provide alarm messages in the event of an issue. The movement of animals in dark and forested environments is tracked using a unique network that is suggested to identify the actions of many animal groups at the same time [10]. The suggested method classifies the animal kind using VGG−19 pre-trained networks, and the Bi-LSTM network generates text-based warning messages using position data. Videos from night vision and security cameras provide both temporal and spatial dynamics. BI-LSTM recurrent networks efficiently manage the temporal details, whereas VGG−19 networks deal with spatial information [11].

Additionally, experimental findings are shown to investigate the legitimate justification outcomes and compare the suggested methodology with previous approaches. The quality of our work is shown by the detailed explanation of the specifics of the many stages of development [12]. Finding the desired outcomes in object detection and classification models is quite difficult. The model performance bottleneck causes poor performance and deteriorates the overall

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

development process in large-scale situations. A greater variety of strategies were used in the previous research to address these situations [13]. Even while the models result in significant increases in accuracy, their performance during testing is subpar.

The following is a summary of the contributions and goals of the suggested techniques: 1) To achieve higher recognition accuracy, the suggested Hybrid VGG-19+Bi-LSTM model is constructed utilising deep neural networks with precisely calibrated hyperparameters.

2) The suggested model uses innovative hybrid techniques in an effort to get exceptional categorisation results.

3) "The suggested system provides foresters with faster SMS alert services and more accurate prediction performance regarding animal detection."

The following is the arrangement of the paper's remaining sections: While Section III outlines the suggested VGG−19+Bi-LSTM system design and implementation specifics, Section II examines comparable studies and points out flaws in earlier advances. The experimental outcomes of the suggested model assessment for four distinct benchmark datasets are shown in Section IV. The description of the whole study and its potential future applications is concluded in Section V.

II. LITERATURE SURVEY

"A survey on scene understanding," S. Chitrakala and S. Aarthi

Scene comprehension has gained significant traction in computer vision in recent years because it can see, analyse, and develop an interpretation of a dynamic scene in real time, leading to new discoveries. A scene is a meaningful perspective of a real-world setting that includes a variety of items and surfaces. While scenes are spread in space and act inside, objects are compact and act upon. Colours, luminosity, contours, shapes, parts, textures, and semantic context are just a few of the qualities that may be used to convey visual information. Making robots seem human is the aim of scene understanding, which aims to fully comprehend visual scenes. With the help of key disciplines like computer vision, cognitive engineering, and software engineering, scene understanding is impacted by cognitive vision. Many prestigious colleges, including Boston University, Stafford Vision Lab, Scene Grammar Lab, Air Lab, Laboratory Machine Vision, and Pattern Recognition, have been steadfastly striving for further advancements in this field as a result of its tremendous expansion. An detailed examination of scene understanding using a variety of techniques and methodologies is included in this study.

"A joint representation of region layout and hierarchy is the connected segmentation tree," S. Todorovic and N. Ahuja,

The connected segmentation tree (CST), a novel object representation proposed in this research, represents the object's canonical qualities in terms of the spatial, geometric, and photometric adjacency and confinement aspects of its individual picture areas. In addition to the recursive embedding structure that is already present in the objectpsilas segmentation tree (ST), CST is created by supplementing the ST with inter-region neighbour linkages. Thus, CST is a region adjacency graph hierarchy. An expansion to regions of the Voronoi diagram for point patterns is used to calculate a regionpsila's neighbours. matched the CST graph representations of unlabelled training pictures and fusing their maximum matched subgraphs is the formula for unsupervised learning of a category's CST model. By concurrently looking for the most salient nodes (areas) and the most salient edges (containment and neighbour links of regions) throughout the picture graphs, a novel learning technique is put forward that

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

maximises the model structure. All instances of the category are simultaneously detected, segmented, and recognised when the category model is matched to the CST of a fresh picture. These outcomes are then semantically explained.

"An empirical investigation into machine learning algorithms for the prediction of heart disease,"

N. K. Kumar, D. Vigneswari, P. K. Rangarajan, and T. A. Assegie,

Machine learning has improved its precision and accuracy in classifying clinical cardiac disease datasets in recent years. Nonetheless, research indicates that the prediction model's performance is significantly impacted by the calibre of the heart disease feature utilised for training. Therefore, using recursive feature elimination with cross-validation (RFECV), this work aims to investigate how the quality of heart disease characteristics affects the machine learning model's performance in heart disease prediction. The research also looks at heart disease characteristics that have a big impact on model results. The University of California Irvine (UCI) machine learning dataset is the source of the experimental dataset. Support vector machines (SVM), logistic regression (LR), decision trees (DT), and random forests (RF) are used in the experiment. the SVM, LR, DT, and RF models' respective performances. The outcome seems to demonstrate that the feature's quality has a major impact on the model's performance. All things considered, the experiment demonstrates that RF performs better than other methods. In conclusion, RF has a 99.7% prediction accuracy.

"Detecting animals with a deep learning algorithm,"

N. Banupriya, S. Saranya, R. Swaminathan, S. Harikumar, and S. Palanisamy, "Vehicular-animal collisions pose a growing risk to both people and wildlife." Observing wild creatures effectively is essential. Both wildlife conservation and the reduction of human-wildlife conflicts need affordable methods for watching the behaviour of wild animals. Therefore, the identification of wild animals requires an effective technique. It is difficult to manually identify every species since there are so many distinct kinds. Tracing animals and preventing accidents involving animals and vehicles are two benefits of animal detection. Effective deep learning techniques will be used to do this. The goal is to develop an algorithm for identifying untamed animals. The model is constructed using the depth-wise separable convolution layer, which blends depth-wise and pointwise convolution. To preserve edge characteristics and control the output image's size, the proposed model inserts zero padding. The suggested method is tested using IWildCam data. With an intersection over union value of 0.878 for detection and 99.6% accuracy for wild animal categorisation, we obtained encouraging findings.

III. SYSTEM ANALYSIS AND DESIGN

EXISTING SYSTEM

Using a cross-frame temporal patch verification method for temporal details and a multi-level graph cut approach for geographical features, the authors Zhang et al. suggested wild animal identification. The model examines the camera trap recordings' background and foreground information. Based on crowded backdrop films, this method segments the animal item from natural situations using a Camera trap and Change Detection net dataset. Despite having a high detection rate, the model struggles to identify important characteristics like location and human interruptions. Convolutional neural networks (CNNs) were suggested by the authors [14] for animal detection, and Iterative Embedded Graph Cut (IEGC) techniques were suggested for creating regions over images, DeepCNN features, and machine learning classification algorithms [15] for classification.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

Even though these models confirm whether the extracted patches are background or animal, their classification performance still needs to be improved.

Deep learning techniques for object detection reached new heights in computer vision applications. Higher assistance for identifying different things in an image or video is provided by the use of object localisation and classification algorithms for object detection. We can determine how many items there are and how active they are based on the extracted findings. Vehicle number plate recognition, traffic monitoring, identifying vehicle theft, tracking objects in hidden boxes, tracking fraudulent activity in public and crowded areas, and Object Character Recognition (OCR) are all applications that heavily rely on this technique for security and video surveillance [16].

The purpose of this paper is to track animal movements inside the forest area, notify forest officials of any hunting or crossing of the forest boundaries, any obstruction to villages or visitors, and any indication of trespassing. The evolution of numerous approaches for using object detection in various settings and applications demonstrates the significance and advancement of object detection in research domains and has drawn increased attention. Furthermore, further studies in this field provide strong frameworks for object detection in various contexts and offer helpful insights into a wide range of applications. Because of their enhanced training performance and excellent accuracy, Fast R-CNN algorithms [17] are often used for object identification. By using region-based networks with whole image-based convolution features, the Faster R-CNN approach [18] quickly enhances the model's detection performance. Compared to previous methods, the Histogram of Orientated Gradients (HOG) feature descriptors [19] locate the objects more quickly by using Region of Interest (ROI) methodologies. By combining ConvNet with region proposal networks, the traditional R-CNN technology [20] presents effective detection techniques. This technique uses annotated information to identify the hundreds of item types included in a picture or video. To anticipate the object areas, the R-CNN algorithms do not use hashing or approximation techniques. Weighted full convolution layers are used in R-FCN approaches [21] to identify an object's area and determine its ROI in order to identify the object's category and background information. With the use of deep learning approaches, object identification methods also seem promising in the domains of traffic scene object recognition [23] and autonomous cars [22].

Using sub-region-based fixed length representations, the Spatial Pyramid Pooling (SPP-net) [25] computes the feature maps in single computations and offers high robustness to the object detection tasks. The Single Short Detector (SSD) methodology [24] employs discretisation techniques based on bounding boxes to efficiently handle feature map information and large volume data. In real-time scenarios, the You Only Look Once (YOLO) architecture processes 155 frames per second, yielding speedier results. This method yields impressive results in object identification with a decreased false-positive rate by using an end-to-end approach to object recognition utilising regression and probabilistic calculations rather than taking categorisation methodologies into consideration. The researchers conduct a thorough analysis of background subtraction and removal. To find the background information, the authors used a variety of techniques, including background cut [28], global statistic-based methods [27], non-parametric models [26], and multiple hypothesis estimation.

Disadvantages

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

• Data complexity: To identify wild animal activity, the majority of machine learning models now in use need to be able to correctly comprehend large and intricate datasets.

• Data availability: In order to provide precise predictions, the majority of machine learning models need a lot of data. The accuracy of the model may degrade if data is not accessible in large enough amounts.

• Inaccurate labelling: The accuracy of the machine learning models that are now in use depends on how well the input dataset was used for training. Inaccurate labelling of the data prevents the model from producing reliable predictions.

PROPOSED SYSTEM

Data pre-processing, animal detection, VGG-19 pre-trained model-based classification, extracting the prediction results, and warning message sending are the five stages of development that make up the suggested architecture. 45k animal photos were gathered during the data pre-processing stage from several datasets, including the camera trap, wild animal, and hoofed animal datasets. The gathered photos were denoised and rescaled to 224 by 224 pixels.

The second step involves feeding the pre-processed photos into the YOLOR object identification model [39], which uses bounding boxes, as shown in Fig. 4, to identify the animal in a picture. The third step involves the use of a hybrid VGG−19+Bi-LSTM model to conduct picture classification tasks, class label prediction, and LSTM Networks to extract animal features. The fourth step involves gathering the animal's position data, after which the web server generates an SMS alert and notifies the forest officials. Lastly, the forest officials will take corrective measures to save the lives of the animals and people.

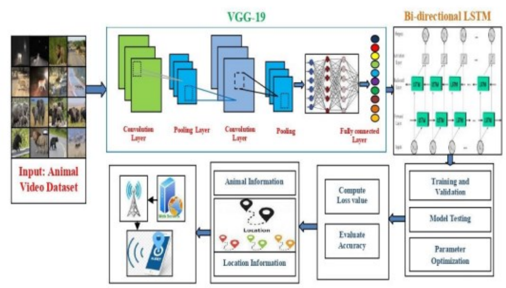
ADVANTAGES

1) To achieve higher recognition accuracy, the suggested Hybrid VGG-19+Bi-LSTM model is constructed utilising deep neural networks with precisely calibrated hyperparameters.

2) The suggested model uses innovative hybrid techniques in an effort to get exceptional categorisation results.

3) "The suggested system provides foresters with faster SMS alert services and more accurate prediction performance regarding animal detection.".

SYSTEM ARCHITECTURE



IV. IMPLEMENTATION

Modules

Service Provider

The Service Provider must use a working user name and password to log in to this module. He may do many tasks after successfully logging in, including Train & Test Data Sets, See the Accuracy of Trained and Tested Datasets in a Bar Chart View Accuracy Results for Trained and Tested Datasets, Download Predicted Data Sets, View Cyber Attack Prediction Status Ratio, and View Cyber Attack Prediction Status See the results of the Cyber Attack Prediction Status Ratio. See Every Remote User.

View and Authorize Users

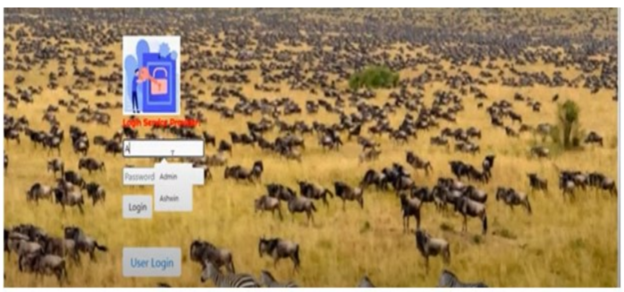
The administrator may see a list of all registered users in this module. Here, the administrator may see the user's information, like name, email, and address, and they can also grant the user permissions.

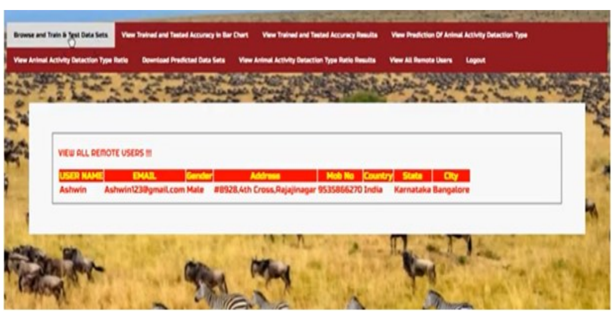
International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

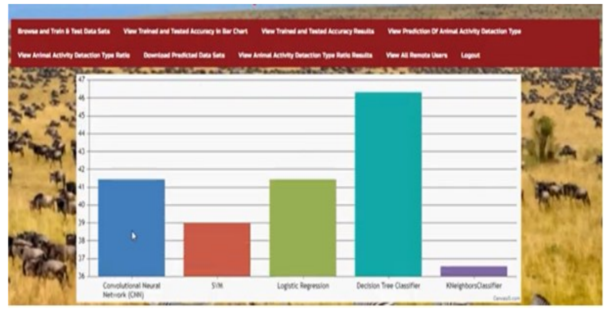
Remote User

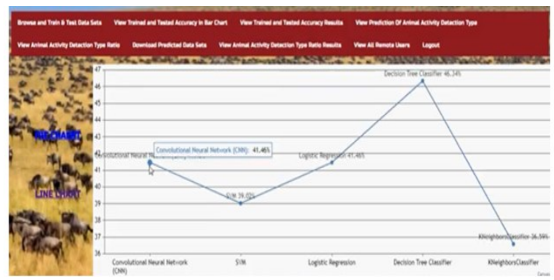
A total of n users are present in this module. Before beginning any actions, the user needs register. Following registration, the user's information will be entered into the database. Following a successful registration, he must use his password and authorised user name to log in. Following a successful login, the user may do tasks including registering and logging in, predicting the status of cyberattacks, and seeing their profile.

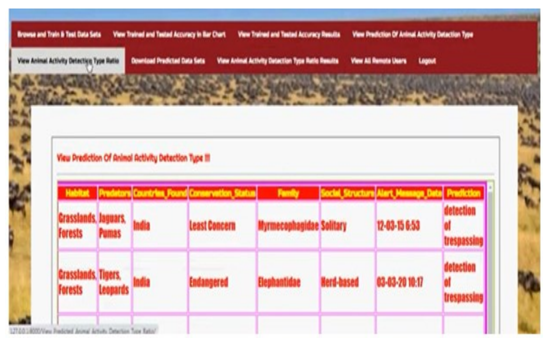
V. SCREEN SHOTS

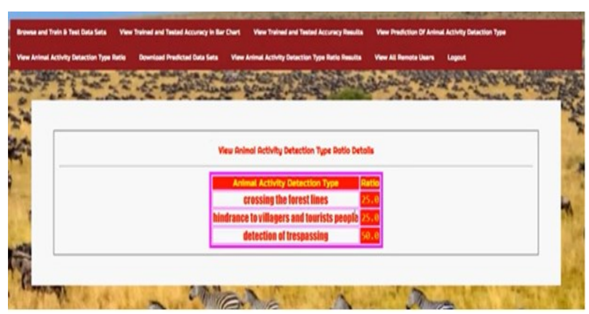










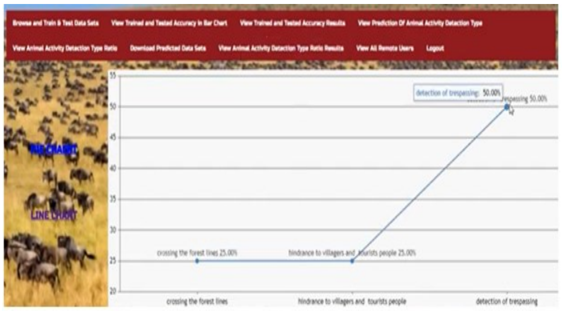


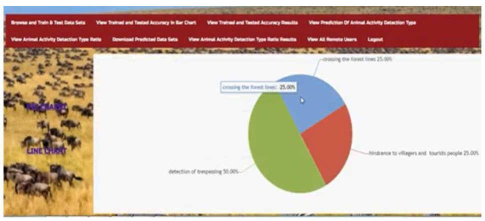
International Journal of Engineering Research and Science & Technology

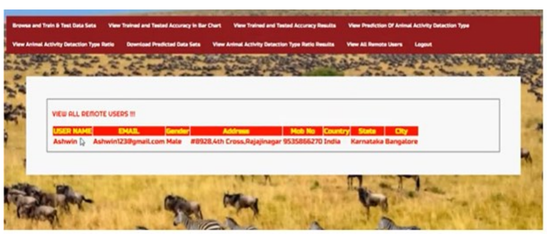
ISSN 2319-5991 www.ijerst.com

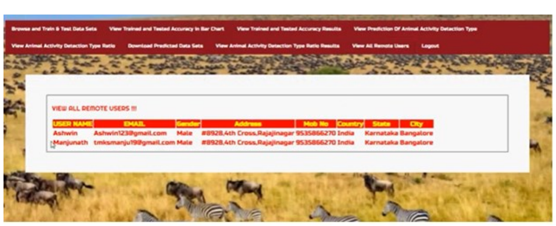
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015

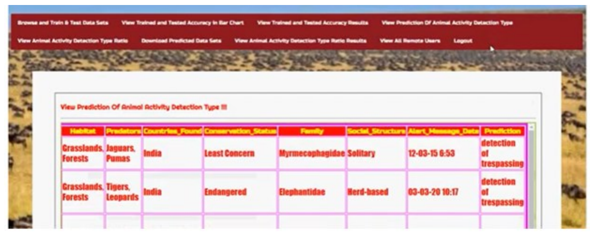
Vol. 21, Issue 2, 2025

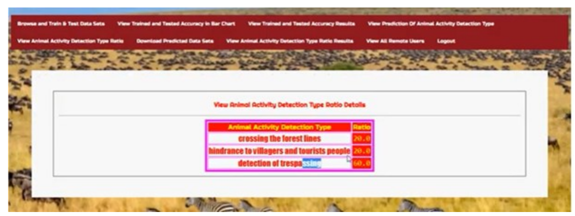


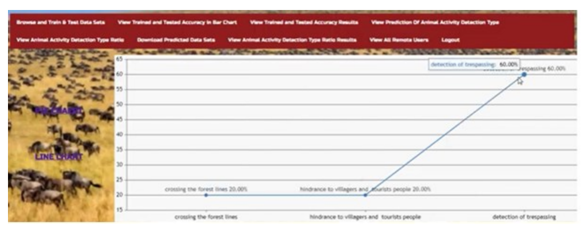


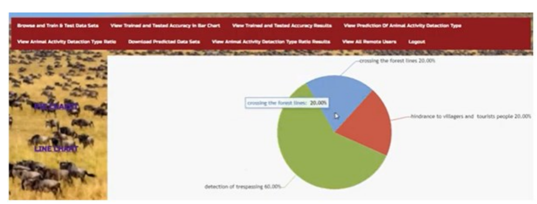












VI. CONCLUSION

The hybrid VGG−19+Bi-LSTM framework for identifying wild animals and tracking their movements is presented in this article. By alerting the forest ranger, this hybrid technique significantly reduces the risk of human hunting and protects people from unexpected animal assaults. This model presents new ways to improve deep learning methods' performance in real-time scenarios and broader applications.

International Journal of Engineering Research and Science & Technology
ISSN 2319-5991 www.ijerst.com
https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015
Vol. 21, Issue 2, 2025

Four distinct benchmark datasets including animal-based data—the camera trap dataset, the wild animal dataset, the hoofed animal dataset, and the CD net data set—have been used to assess the suggested approach. The experimental findings demonstrate our model's enhanced performance across a range of quality parameters. Above 98% average classification accuracy, 77.2% mean Average Precision (MAP), and 170 FPS are achieved using the suggested hybrid VGG-19+Bi-LSTM model. The suggested hybrid VGG-19+Bi-LSTM model now outperforms previous methods and yields better results in less computing time.

REFERENCES

[1] S. Aarthi and S. Chitrakala, ''Scene understanding—A survey,'' in *Proc. Int. Conf. Comput., Commun. Signal Process. (ICCCSP)*, Jan. 2017, pp. 1–4.

[2] N. Ahuja and S. Todorovic, ''Connected segmentation tree—A joint representation of region layout and hierarchy,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8. [3] 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.

[4] 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.

[5] 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.

[6] J. Dai, Y. Li, K. He, and J. Sun, ''R-FCN: Object detection via region-based fully convolutional networks,'' 2016, *arXiv:1605.06409*.

[7] 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.

[8] B. Natarajan, E. Rajalakshmi, R. Elakkiya, K. Kotecha, A. Abraham, L. A. Gabralla, and V. Subramaniyaswamy, ''Development of an end-to end deep learning framework for sign language recognition, translation, and video generation,'' *IEEE Access*, vol. 10, pp. 104358–104374, 2022.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com

https://doi.org/10.62643/ijerst.2025.v21.i2.pp1006-1015

Vol. 21, Issue 2, 2025

[15] G. Farnebäck, ''Two-frame motion estimation based on polynomial expansion,'' in *Proc. 13th Scandin. Conf. (SCIA)*. Halmstad, Sweden: Springer, Jul. 2003, pp. 363–370.

[16] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[17] 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.

[18] 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.

[19] K. He, X. Zhang, S. Ren, and J. Sun, ''Spatial pyramid pooling in deepconvolutional networks for visual recognition,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.

[20] 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.

[21] 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.

[22] 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.

[23] T. Liang, H. Bao, W. Pan, X. Fan, and H. Li, ''Detect Former: Category assisted transformer for traffic scene object detection,'' *Sensors*, vol. 22, no. 13, p. 4833, Jun. 2022.

[24] G. Li and Y. Yu, ''Deep contrast learning for salient object detection,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 478–487.

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in *Computer Vision—ECCV*. Amsterdam, The Netherlands: Springer, 2016, pp. 21–37.