

Safeguarding Wildlife With Advanced Neural Networks and Real-time Alerts For Wild Animal Activity Detection

^{1.} **R. Harshitha**, B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY, rambhaharshitha2002@gmail.com

^{2.} **I. Keerthi**, B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY, Keeethiinje7@gmail.com

^{3.} **T. Neveen Kumar**, B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY, tnddytanukula@gmail.com

^{4.} **B. D V Siva Sai**, B.Tech, Department of CSE, DNR COLLEGE OF ENGINEERING AND TECHNOLOGY, bdvsiva018@gmail.com

^{5.} **Mrs. V. Navya Devi**, M. Tech, Assistant Professor, Department of computer Science and Engineering, navyadevidnr@gmail.com

ABSTRACT: The creation of an effective monitoring model is the primary goal of this study, which seeks to alleviate the serious problem of animal assaults experienced by forestry workers and rural residents. In order to enhance safety warnings in woodland regions, the proposed network combines the Hybrid Visual Geometry Group (VGG)-19 with Bidirectional Long Short-Term Memory (Bi-LSTM). Its goals include animal type detection, movement monitoring, and the provision of up-to-the-minute position information. The model outperforms conventional techniques of surveillance by accurately detecting animals and their movement patterns with the integration of VGG-19 for feature extraction and Bi-LSTM for sequence learning. Additionally, an ensemble approach is used to improve accuracy and resilience by aggregating predictions from numerous separate models. To get perfect accuracy, it is recommended to investigate methods like as CNN+BiGRU, which significantly improve performance. An enhancement to the implementation is a Flask-built, user-friendly front end that allows for authentication feature testing during user testing. By combining state-of-the-art deep learning methods with user-centric design, this study presents a viable option for improving safety monitoring in rural and forested areas, therefore reducing the likelihood of animal attacks. VGG-Net, Bi-LSTM, convolutional neural networks, activity recognition, video surveillance, monitoring of wild animals, warning system are all terms that are used in the context of animal identification.

I. INTRODUCTION:

The increasing number of human-wildlife encounters in recent years has highlighted the urgent need for new ways to reduce conflicts and make everyone's lives safer. Conflicts and dangers between people and animals have increased due to habitat encroachment, which is caused by the growth of cities and farms [1]. The development of efficient early warning systems is crucial for conservationists, lawmakers, and local populations to address these encounters in a timely manner by mitigating dangers and facilitating responses. Concerns about human-wildlife confrontations have recently gained traction, and one potential answer is the Wild Animal Activity Alerting System (WAAAS). The goal of WAAAS is to use state-of-the-art ML and DL algorithms to identify and alert humans when wild animals are in close proximity to human areas. This will help with proactive management tactics [2]. Waaas aims to provide accurate and timely warnings by using data from several sources, including photos, motion sensors, and sound recordings, to determine patterns that indicate the presence and behavior of animals [3].

Rising tensions between people and other forms of life, brought driven by factors like habitat loss and competition for few resources, highlight the critical need for such a system. Herbivore crop raids, animal predation, and even human assaults are all ways in which these conflicts show themselves. In particular, these occurrences endanger animal populations and derail conservation initiatives, in addition to endangering human lives and livelihoods [4]. As a result, more and more people are realizing that we need to do something to help animals and humans live together peacefully [5].

In this light, the creation of WAAAS is a giant leap forward in terms of improving security, reducing economic losses, and furthering conservation goals. Waaas aims to analyze signs of wild animal behavior with unparalleled precision and efficiency by applying state-of-the-art ML and DL algorithms [6]. This introduction explores the complex issues surrounding human-wildlife interactions, highlights the role of early warning systems in finding solutions, and outlines the goals and reach of WAAAS in reducing conflicts and promoting harmony.

A Rising Anxiety Driven by factors such as habitat fragmentation, climate change, and increasing human populations, the intensifying contact between people and animals has become a significant worldwide problem [7]. Human activities are encroaching into animal territories and natural habitats are decreasing, leading to an increase in human-wildlife conflicts across many ecosystems [8]. These confrontations have far-reaching consequences, impacting people in both rural and urban settings. In rural areas, crops are damaged and animals are preyed upon. In urban areas, wildlife encroachment and property damage are prevalent problems. Herbivores like wild boars, elephants, and deer may wreak havoc on agricultural livelihoods, causing terrible economic losses and making food insecurity even worse for already-vulnerable populations [10]. Similarly, conservationists and pastoralists sometimes find themselves at odds when animals such as wolves, lions, and bears prey on sheep. This may lead to retaliatory murders and worsen the conservation issues already faced [11]. Additionally, local populations and wildlife authorities experience heightened tensions when huge predators sometimes attack people, which causes anxiety and insecurity [12].

Because animal behavior is always changing and there are so many different species involved, efforts to reduce human-wildlife interactions are already challenging. Different species need different approaches to management because to their distinct feeding habits, territorial behaviors, and reactions to human interference [13]. Also, human-wildlife interactions may vary in both space and time, which is why early warning systems and real-time monitoring are crucial for facilitating preemptive interventions [14].

II. LITERATURE SURVEY

Research into creating reliable techniques for comprehending complicated visual situations has been vigorously pursued due to the growing interest in scene comprehension. In their extensive review of scene understanding strategies, Aarhi and Chitrakala [1] show how many different ways people in this field tackle the problem. Robotics, autonomous navigation, and surveillance are only a few of the many areas where scene understanding is crucial, as shown by their work.

One effective way to depict the structure and hierarchy of an image's regions is using connected segmentation trees (CSTs). The idea of CSTs is introduced by Ahuja and Todorovic [2]. CSTs allow for the combined representation of area layout and hierarchy, which helps with fast object detection and picture segmentation. Among the many fields that have made use of this method are scene analysis, medical imaging, and remote sensing. Disease prediction is one area where machine learning algorithms have found extensive use in predictive modeling. Evaluating the effectiveness of several classifiers on a heart illness dataset, Assegie et al. [3] provide an empirical research on machine learning techniques for heart disease prediction. Their research provides valuable information for clinical decision-making and risk stratification by illuminating the effectiveness of several algorithms in predicting the likelihood of cardiovascular disease.

With the advent of deep learning techniques, computer vision has seen a sea change, with previously unimaginable improvements in object identification and detection. Investigating the use of deep learning methods for animal recognition, Banupriya et al. [4] show that convolutional neural networks (CNNs) are effective in picture animal identification and classification. Their research demonstrates how deep learning methods might be useful for conservation and animal monitoring.

In object detection tasks, objectness estimate is essential for finding areas that are likely to contain the items of interest. Binarized Normed Gradients (BING) is an objectness estimate approach proposed by Cheng et al. [5] that reaches 300 fps real-time performance. A lightweight and effective approach for object recognition in photos is offered by BING, which leverages binarized features and normed gradients. Modern object identification methods use region-based techniques for precise localization and classification; one such method is region-based fully convolutional networks (R-FCNs). Dai et al. [6] provide R-FCN, a method for efficient object recognition in pictures that combines fully convolutional networks (FCNs) with region proposal networks (RPNs). Based on their findings, R-FCN can successfully outperform its competitors on benchmark datasets.

Applications such as environmental monitoring, urban planning, and surveillance rely heavily on change detection. A change detection framework based on weightless neural networks is proposed by De Gregorio and Giordano [7]. These networks are resilient to noise and can adapt to dynamic settings. Applications in land cover mapping and disaster management are possible thanks to their method, which shows promise in identifying changes in remote sensing pictures.

The ever-changing and intricate nature of sign motions makes sign language identification a formidable challenge. In their paper, Natarajan et al. [8] provide a comprehensive deep learning system that can recognize signs, translate them, and even create videos. Their study shows promise as a solution to help the deaf communicate by using deep neural networks to extract information from films of sign language and accurately recognize and translate sign motions.

III. METHODOLOGY

a) Proposed work:

b) An animal activity detection network that combines VGG-19 and Bi-LSTM will be created and tested as part of the proposed study. Improved safety via timely alarms is achieved by the integration of feature extraction and sequence learning in this model, which enhances accuracy and enables real-time monitoring. To further improve accuracy, the project may be extended to include a CNN+GRU model. This model combines a Bidirectional GRU layer with the CNN[16] algorithm. GRU was used because it outperformed LSTM in optimizing picture features. On top of that, we have built a Flask framework with SQLite to make user registration and login easier. This will allow users to try out the system's features and make it more user-friendly. These updates are an attempt to fix the problems with animal activity detection in forests in a more solid and efficient way.

(b) Network Design:

The input dataset, which comprises of photos taken in forest zones, is the starting point for the system design. Following normalization and augmentation as part of the preprocessing, these pictures are divided into two sets: one for training and one for testing. Three distinct algorithms, each designed for animal activity detection: CNN, VGG19-BiLSTM, and CNN+Bidirectional GRU, are used during the training phase. To extract features, the CNN model employs convolutional layers; to learn sequences, the VGG19-BiLSTM combines convolutional layers with Bi-LSTM. To get better results, the CNN+Bidirectional GRU model integrates CNN with a Bidirectional GRU layer. After the models have been trained, they are tested on a different set of data to see how well they can identify animal behavior. Enhancing safety in forest situations, the detection model evaluates input photos and delivers timely notifications in the case of animal activity. It is capable of real-time monitoring.

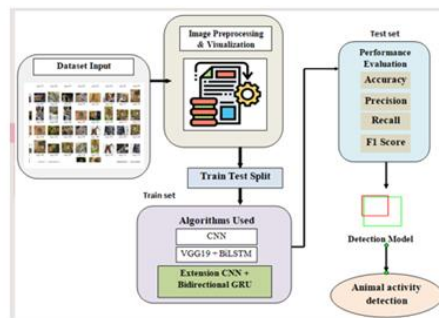


Fig 1 Proposed Architecture

c) Dataset collection:

The camera trap dataset [48], the wild animal dataset [49], the hoofed animal dataset [50], and the CDnet dataset [51] were the four separate benchmark datasets whose photos were acquired throughout the data set gathering procedure. These datasets contain a wide variety of habitats and animal types, making them ideal for training and testing the suggested model. The wild animal dataset provides a more diverse range of species and behaviors, while the camera trap dataset provides images of animals in their natural environments. The hoofed animal dataset also adds species-specific features to the training data by concentrating on hoofed animals. By making available annotated video sequences for assessment, the CDnet dataset adds to the variety. The camera trap dataset [48], the wild animal dataset [49], the hoofed animal dataset [50], and the CDnet dataset [51] were the four separate benchmark datasets whose photos were acquired throughout the data set gathering procedure.

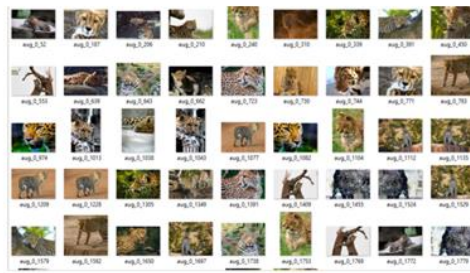


Fig 2 DATA SET

These datasets contain a wide variety of habitats and animal types, making them ideal for training and testing the suggested model. The wild animal dataset provides a more diverse range of species and behaviors, while the camera trap dataset provides images of animals in their natural environments. The hoofed animal dataset also adds species-specific features to the training data by concentrating on hoofed animals. By making available annotated video sequences for assessment, the CDnet dataset adds to the variety. Effective identification of wild animal activity over a wide range of environmental circumstances and species compositions is made possible by the model's incorporation of several datasets, which provide a strong and diversified training corpus.



Fig 3 DATA SET

d) **DATA PROCESSING**

For deep neural networks to identify activities in wild animals, there are a number of critical stages involved in the dataset preparation stage. To provide uniformity across the dataset and easier convergence during training, the pictures are first normalized to standardize their pixel values. To further improve the model's generalizability and decrease bias, the photos are then randomly shuffled to change their order. To further improve the dataset's variety and resilience, augmentation methods like flipping, rotating, or cropping may be used. Plus, we mark the photos to show if there are any wild animals in them or not. The preprocessing step lays the groundwork for successful training and assessment of the hybrid deep neural networks by methodically normalizing, shuffling, and labeling the dataset. This helps make sure the models can identify and categorize wild animal behavior in different kinds of environments, which in turn helps with making sure people in wildlife areas get alerts when they need them.

f. Envisioning

Dataset insights and model performance evaluation rely heavily on visualization using Seaborn and Matplotlib. Histograms, scatter plots, heatmaps, and many other tools for making informative charts and plots are available in these libraries. Users may effortlessly create visually beautiful visualizations with minimum code using Seaborn's high-level interface, which is built on top of Matplotlib. These visuals are useful for seeing trends that could affect the training and performance of models, finding outliers, and comprehending data distribution. It is possible to evaluate the performance and behavior of the model in its entirety by viewing its metrics, including loss, confusion, and accuracy matrices. Using Seaborn and Matplotlib, researchers may improve the analysis and models that come out of it by efficiently communicating results, validating hypotheses, and making informed choices all along the machine learning pipeline.

g) Extraction of Features

An essential step in detecting the behavior of wild animals using hybrid deep neural networks is feature extraction. To extract features that differentiate between various animal activities, this context uses image analysis to find relevant patterns and traits in the input photographs. The capacity of convolutional neural networks (CNNs) to automatically build hierarchical representations from raw pixel input makes them a popular choice for feature extraction. A convolutional neural network (CNN) and a recurrent neural network (RNN), such as an LSTM or a GRU, work together to extract features in the suggested hybrid model. Using the input pictures, the convolutional neural network (CNN) component extracts spatial features, while the recurrent neural network (RNN) component analyzes time sequences to extract context and motion. This hybrid model successfully extracts geographical and temporal characteristics by combining the two kinds of neural networks. It can then accurately identify wild animal behavior and provide alarm messages to notify the appropriate authorities or persons.

h) Evaluation/Training

To build a reliable system that can identify the activities of wild animals utilizing hybrid deep neural networks and generate alarm signals, testing and training are crucial steps. The hybrid model sees a tagged dataset of pictures showing different kinds of wild animal behavior as it's being trained. The model learns to extract useful characteristics and provide reliable predictions using backpropagation and other iterative parameter optimization approaches. As part of the training process, we check the model's accuracy on a portion of the data to make sure it can handle new situations.

Testing, on the other hand, involves mimicking real-world situations by evaluating the trained model on a separate set of data. Metrics including F1 score, recall, accuracy, and precision are used to evaluate the model's capacity to reliably identify wild animal activities and send out alarm signals. Thanks to testing, researchers may see how well the model works, find any problems it may have, and fix them by improving the model's architecture or training procedure. Developing a dependable system that may successfully mitigate dangers connected with interactions with wild animals requires extensive testing and training.

i) Presenting algorithms CNN

The deep learning architecture known as a Convolutional Neural Network (CNN) was developed with the express purpose of performing image identification tasks. To identify when wild animals are active, our study makes use of an existing CNN[16]. Convolutional neural networks (CNNs) include feature extraction stages in the form of convolutional layers, dimensionality reduction stages in the form of pooling layers, and classification stages in the form of fully connected layers. The research makes use of convolutional neural networks (CNNs) to analyze raw visual data and extract pertinent elements that indicate the presence of wild animals. This model is able to successfully detect animal movement by using the CNN's [16] hierarchical feature extraction capabilities to identify patterns in pictures, categorize them, and then create alarm signals.

Vote for VGG19 + BI-LSTM

In the VGG19 + BI-LSTM model, which has been suggested, the VGG19 CNN[17] architecture is combined with the BI-LSTM recurrent neural networks. In order to capture temporal relationships, BI-LSTM analyzes sequential information, while VGG19 extracts hierarchical features from input pictures. The research makes use of this hybrid model to identify the presence of wild animals. BI-LSTM examines sequences of feature vectors across time to identify patterns that suggest changes in animal behavior, while VGG19 collects spatial information from pictures. The VGG19 + BI-LSTM model improves the accuracy of detecting wild animal activity and allows for the timely creation of alarm signals by integrating spatial and temporal information.

NBC and GRU

To create the CNN + GRU model, we combined two popular models: CNNs[17] and Gated Recurrent Units. GRUs analyze sequential data to identify temporal relationships, while CNNs extract spatial properties from pictures. The research makes use of this hybrid model to identify the presence of wild animals. The convolutional neural network (CNN) part of the system takes in pictures and uses them to extract spatial features; the gradient recurrent unit (GRU) part uses feature vector sequences across time to find patterns that might indicate changes in the animals' behavior. The CNN + GRU[17] model improves the accuracy of detecting wild animal activity by combining geographical and temporal information, which allows for the timely creation of alarm signals.

IV. EXPERIMENTAL RESULTS

Accuracy: A test's accuracy is defined by how well it distinguishes between healthy and sick samples. We can determine a test's accuracy by calculating the percentage of reviewed instances with true positives and true negatives. It is possible to express this mathematically as:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

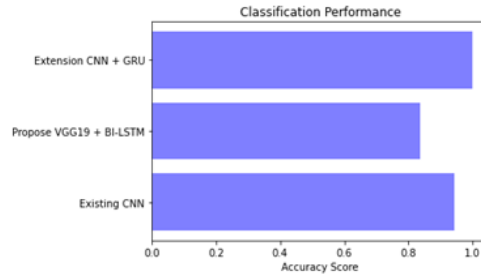


Fig 4 ACCURACY COMPARISON GRAPHS

Precision: The accuracy rate, or precision, is the percentage of true positives relative to the total number of occurrences or samples. Consequently, the following is the formula for determining the accuracy: Precision is TP divided by (TP plus FP), which is the sum of true positives and false positives.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

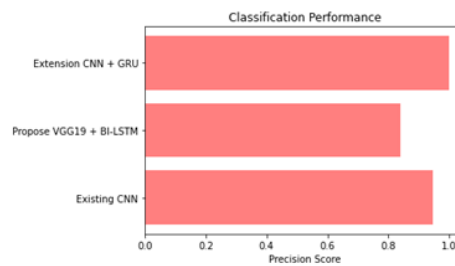


Fig 5 PRECISION COMPARISON GRAPHS

Recall: Recall is a machine learning statistic that evaluates a model's capacity to detect all significant occurrences of a given class. The completeness of a model in capturing instances of a particular class is shown by the ratio of properly predicted positive observations to the total actual positives.

$$\text{Recall} = \frac{TP}{TP + FN}$$

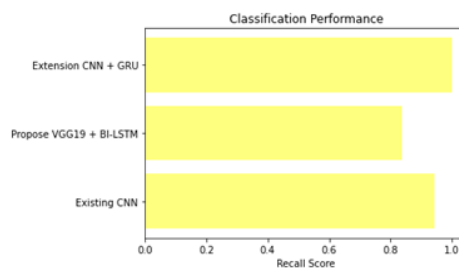


Fig 6 RECALL COMPARISON GRAPHS

F1-Score: One way to evaluate a model's performance in machine learning is via its F1 score. This method integrates a model's recall and accuracy scores. A model's accuracy may be measured by counting the number of times it correctly predicted something throughout the whole dataset.

$$\mathbf{F1\ Score} = \frac{2}{\left(\frac{1}{\mathbf{Precision}} + \frac{1}{\mathbf{Recall}}\right)}$$

$$\mathbf{F1\ Score} = \frac{2 \times \mathbf{Precision} \times \mathbf{Recall}}{\mathbf{Precision} + \mathbf{Recall}}$$

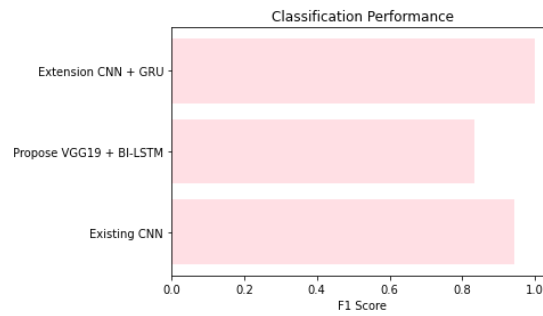


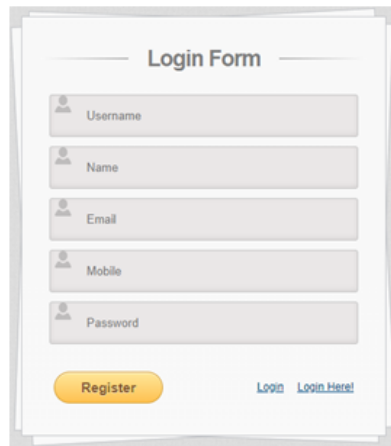
Fig 7 F1 COMPARISON GRAPHS

	MLModel	Accuracy	Precision	Recall	f1_score
0	Existing CNN	0.943	0.946	0.943	0.943
1	Propose VGG19 + BI-LSTM	0.837	0.841	0.837	0.834
2	Extension CNN + GRU	1.000	1.000	1.000	1.000

Fi 8 table



Fig 9 HOME PAGE



The image shows a registration form titled "Login Form". It contains five input fields: "Username", "Name", "Email", "Mobile", and "Password". Below the fields is a yellow "Register" button and a blue link that says "Login Login Here!".

Fig 10 SIGN UP



The image shows a login form titled "Login Form". It has a text input field with the value "admin" and a password field with masked characters "*****". Below the fields is a yellow "Login" button and a blue link that says "Register Register Here!".

Fig 11 SIGN IN

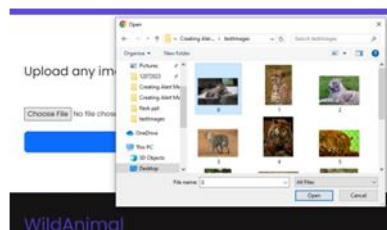


Fig 12 upload input data

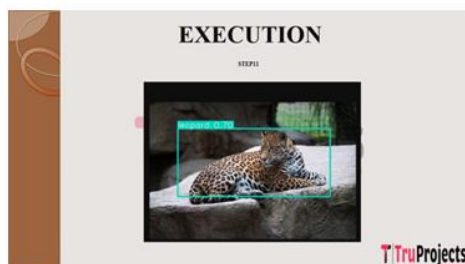


Fig 13 predicted result

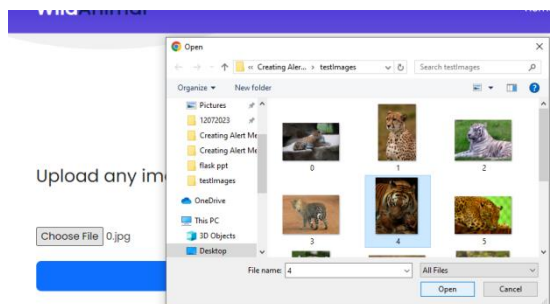


Fig 14 upload input data



Fig 15 predicted result

V. CONCLUSION

Finally, exceptional performance in classifying the activities of wild animals has been achieved by the creation and integration of CNN, VGG19 + BI-LSTM, and CNN-GRU algorithms. We have minimized computing costs via simplified methods, guaranteeing that our forest monitoring operations will be sustainable. Hybrid CNN-GRU models have shown better detection robustness than separate models by combining the best features of both CNN and GRU. Moreover, the system's usability is improved by integrating a Flask front-end with a SQLite database. This allows for simple data entry and display of animal detection results. By developing a reliable, user-friendly, and widely disseminated system for tracking the movements of wild animals, this initiative will ultimately help forestry professionals, conservationists, and residents of rural areas. This initiative helps people and animals live in harmony by making forests safer and increasing conservation efforts, which is great for the environment.

VI. FUTURE SCOPE

A number of critical components make up the warning message production system that is based on the detection of wild animal behavior utilizing hybrid deep neural networks. To begin with, the system is designed to identify and categorize various patterns and behaviors shown by animals in their natural habitats. These patterns and behaviors may include movements, interactions, and any unusual occurrences. The technology also aims to examine animal activity sequences across time in order to spot patterns that might be signs of danger or distress. Optimization of computing expenses and sustainability of forest monitoring activities are also within the system's feature scope. It also involves making the interface easy to use so that data can be submitted quickly and the results of animal detection can be shown. The overarching goal of this feature is to promote peaceful cohabitation between people and animals by increasing safety and conservation efforts in wooded areas by the prompt alerting of important stakeholders, such as rural communities, forestry workers, and conservationists.

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, "RFCN: 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. Subramaniaswamy, "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. Karuppiyah, "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. Subramaniaswamy, 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.
- [15]. G. Farneback, "Two-frame motion estimation based on polynomial expansion," in *Proc. 13th Scand. 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, "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 deep convolutional 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, "DetectFormer: 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.
- [26]. N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 831–843, Aug. 2000.
- [27]. M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.

- [28]. W. Ouyang and X. Wang, “Joint deep learning for pedestrian detection,” in Proc. IEEE Int. Conf. Comput. Vis., Dec. 2013, pp. 2056–2063.
- [29]. M. Monnet, Paragios, and V. Ramesh, “Background modeling and subtraction of dynamic scenes,” in Proc. 9th IEEE Int. Conf. Comput. Vis., 2003, pp. 1305–1312.
- [30]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 779–788.