JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

REAL TIME OBJECT DETECTION USING YOLO

Ms. T .Naga Navya[1] Vunnam Harshini[2], Tadapaneni Divya Prasanna[3], Shaik Farjana[4], Malleboyina Deepthi[5]

[1*]Assistant Professor:Vignan's Nirula Institute of Technology and Science for Women

[2345]B.Tech Scholar:Vignan's Nirula Institute of Technology and Science for Women

tnavyab@gmail.com , Vunnamharshini0@gmail.com ,tadapanenidivya2@gmail.com, farju.sk167@gmail.com, malleboyina @gmail.com

ABSTRACT

Real-time object detection is very useful component in real life applications. We use this detection application in many forms .For example using in Automobiles like Autonomous vehicles, Surveillance Systems and robots .These type of applications useful in enhancing the speed of objects in Dynamic environment .To detect them we use Machine learning Algorithms Particularly CNNs .With the help of this Mechanism we can detect the multiple real time applications.to perform any real time object detection methods we need hardware acceleration which further contribute to the ultimate efficiency of the systems. Now a days Real-time object applications has become increasingly vital Across numerous domains .Algorithms we use to prepare these type of Applications provide in depth analysis of state of the art methodologies, including both Traditional machine learning and contemporary machine learning technique Experimental results paving the way for practical Real-time applications .Every model we prepare to detect must be checked before Applying in real life ,we check that by giving numerous images ,text .Finally by training them for the long period object detection application should be able to give Accuracy values.

KEYWORDS

Real-time object Detection, Convolutional neural network (CNN), Machine learning, Training the object, Optimization.

1.INTRODUCTION:

Object detection is a technique used in computer vision for the identification and localization of objects within an image or a video [1-3]. Image Localization is the process of identifying the correct location of one or multiple objects using bounding boxes, which correspond to rectangular shapes around the objects [4-5]. This process is sometimes confused with image classification or image recognition, which aims to predict the class of an image or an object within an image into one of the categories or classes [6-9].

Real-time object detection is a critical task in computer vision, involving the identification and localization of objects in or video frames as they are captured or streamed [10-12]. Unlike traditional object detection, which may allow for significant processing time, real-time detection demands immediate analysis, typically within milliseconds, to provide instant feedback or decisions [13-15]. Real-time object detection is widely used in applications such as autonomous vehicles, security surveillance, robotics, augmented reality, and smart devices [16-18]. The primary goal is to balance accuracy and speed, enabling systems to recognize multiple objects while maintaining performance [19-20]. Several techniques and algorithms are employed for real-time object detection [21-22]. Early methods relied on handcrafted features and traditional machine learning models like the Viola-Jones detector for face detection [23-25]. We use many Algorithms to perform Real-time object detection For Examples YOLO, SSD, CNN [26-27]. In the following research we performed YOLO to predict the object detection. YOLO is known for its good speed and accuracy ,it can process the whole image at once so we choosed this Algorithm to predict real time applications [28-31].

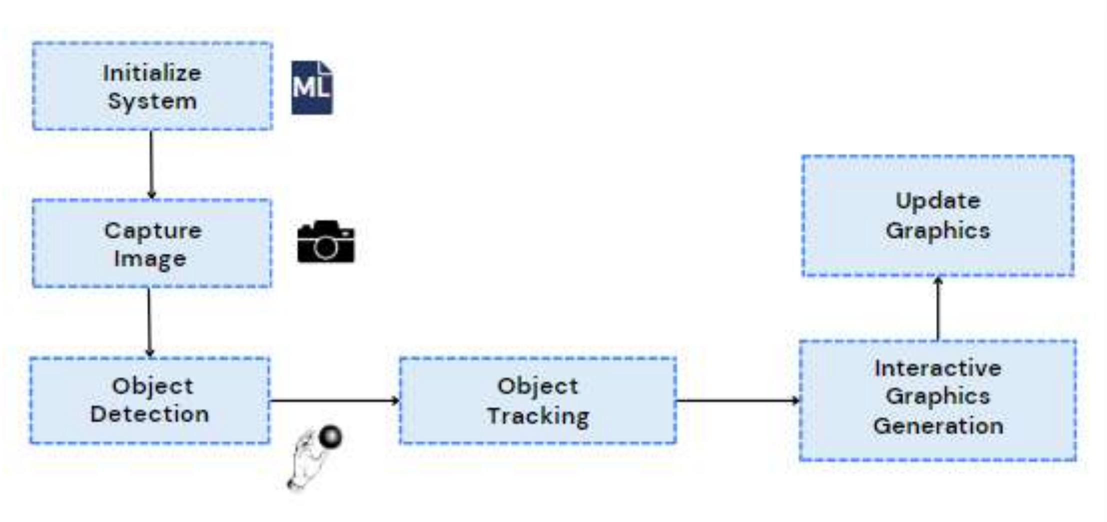

Fig. 1 Block Diagram

2.LITERATURE AND SURVEY

The Literature survey provides some research related to real-time object detection which has gained significant attention in various applications, including surveillance, autonomous vehicles, and robotics. This survey explores the advancements in machine learning techniques, specifically focusing on deep learning, which have revolutionized this field.

Object detection, a critical component of computer vision, has undergone significant advancement, particularly with the emergence of deep learning methodologies. At its core, this field focuses on accurately identifying and pinpointing objects within digital imagery. This capability is crucial for a wide range of applications that impact both our daily life and various industries. For instance, it plays a pivotal role in improving the safety features of self-driving cars and is instrumental in transforming the landscape of medical diagnostic procedures. The application of object detection algorithms enables machines to analyse and comprehend visual data in a manner akin to human perception, thus broadening the scope of what machines can achieve in understanding and interacting with the environment around them. This evolution in object detection not only showcases the power of modern technology but also opens up new possibilities for innovation and efficiency across diverse sectors.

Johannes Schoning [32-34], **Antonio Kruger** [35-36], **Michael Rosh** [37-39], **Markus Lochtefeld** [40-41] had studied on LittleProjectedPlanet: An Augmented Reality Game for Camera Projector Phones. This project explores the integration of miniaturized projection technology, specifically Pico projectors, into mobile devices. This integration enables these devices to project large-scale information on real-world surfaces, expanding the interaction space of mobile devices to physical objects in the environment. This advancement opens up possibilities for innovative interaction concepts that are not achievable on traditional desktop computers. The paper discusses the potential applications of camera projector phones through a mobile adaptation of the PlayStation 3 game LittleBigPlanet. The camera projector unit augments users' hand drawings by overlaying virtual objects and simulating their physical interactions with the real world. Players can create a 2D world on paper or use existing physical objects, allowing the physics engine to simulate interactions and achieve game goals. This research showcases the transformative potential of integrating projection technology into mobile devices paving the way for novel interactive experiences.

Alessandro Dal Corso [42], **Gudmundur Einarsson** [43], had studied on Virtual Table: a projection augmented reality game. In the Virtual Table project, an interactive tower defence game is created using projection augmented reality. Players engage in the game by preventing virtual stylized soot balls from reaching the cheese on a table. Any object placed on the table becomes part of the game, serving as a wall, obstacle, or tower, depending on its type. The game encourages strategic thinking and collaboration, rather than relying solely on the quantity of objects. This continuous installation allows any number of players to join for an optional period, fostering an interactive and engaging experience for participants.

In computer vision, object tracking is a challenging area that researchers need to focus on. To overcome this problem, **B. Drayer** [44], **Yilmaz et al.** [45] reviewed state-of-the-art tracking methods and new trends associated with them. The survey categorized the tracking methods according to the objects and how the object's motion is represented. It described the representative methods in each category and determined the merits and demerits of each method.

William T. Freeman [45], **Girshick** [46-47] had studied Computer Vision for Interactive Computer Graphics. In this research, various vision algorithms were developed and applied for interactive graphics applications. The study focused on fundamental visual measurements, including large object tracking, shape recognition, motion analysis, and small object tracking. These measurements were utilized to create vision-based interfaces for computer games and interactive devices like toy robots and televisions. The research team also designed a specialized image detector/processor, reducing implementation costs. This work demonstrates the practical applications of computer vision in enhancing user experiences and interactive interfaces, showcasing the potential for cost-effective and immersive interactions in diverse contexts.

3.PROPOSAL MECHANISM

3.1 Data Collection

First, you need to gather a dataset that includes the images and the corresponding annotations for the objects you want to detect. This is one of the most important steps in training your YOLOv10 model. Although it might seem tedious and time-consuming, having a high-quality dataset is essential for achieving good performance.

Here are some commonly used tools for creating annotations:

- **LabelIMG**: An easy-to-use graphical image annotation tool that generates XML files in PASCAL VOC format.
- **Labelme**: Another popular annotation tool that outputs JSON files compatible with various frameworks.

3.2 YOLO (You Only Look Once)

YOLO (You Only Look Once) is a state-of-the-art model to detect objects in an image or a video very precisely and accurately with very high accuracy. In this tutorial, we will learn to run Object Detection with YOLO and plot the frames using OpenCV on both a recorded video and a camera.

The YOLO algorithm was introduced by Joseph Redmon et al. in 2015 and has undergone several improvements, leading to various versions such as YOLOv2, YOLOv3, YOLOv4, and YOLOv5. Each iteration has brought enhancements in accuracy, speed, and the ability to detect smaller objects in more complex scenarios

3.3 Preprocessing

Before training, it's essential to preprocess your data. This process involves resizing images, normalizing pixel values, and converting annotations into the format required by YOLOv10. Proper preprocessing is crucial as it ensures the model can learn effectively from the data. Additionally, researching more about data augmentation techniques and potential issues with image preprocessing can be beneficial.

While the YOLO framework includes many preprocessing steps during training, these can be handled automatically, especially if you are new to this process. However, understanding and manually performing these steps can give you better control over the quality and variability of your dataset, ultimately leading to a more robust model.

The YOLO annotation format is:

```
<object-class> <x_center> <y_center> <width> <height>
```

3.4 Importing Necessary Libraries

- PyTorch or TensorFow
- Darknet
- OpenCV
- Numpy
- Matplotlib
- Pillow
- Scipy
- PyYAML
- Tqdm

3.5 ALGORITHM

The basic idea behind YOLO is to divide the input image into a grid of cells and, for each cell, predict the probability of the presence of an object and the bounding box coordinates of the object. The process of YOLO can be broken down into several steps:

1. Input image is passed through a CNN to extract features from the image.
2. The features are then passed through a series of fully connected layers, which predict class probabilities and bounding box coordinates.
3. The image is divided into a grid of cells, and each cell is responsible for predicting a set of bounding boxes and class probabilities.
4. The output of the network is a set of bounding boxes and class probabilities for each cell.
5. The bounding boxes are then filtered using a post-processing algorithm called non-max suppression to remove overlapping boxes and choose the box with the highest probability.
6. The final output is a set of predicted bounding boxes and class labels for each object in the image.

3.6 Loading the YOLO Model and Class Names

1. Import the required libraries: Start by importing the necessary libraries, including YOLO from the Ultralytics package, cv2 (OpenCV), and math.

2. Set up the YOLO model: Initialize an instance of the YOLO class and load the YOLO weights using the path to the weights file you downloaded earlier.

3. Define the class names: Create a list of class names corresponding to the output classes of the YOLO model. The class names are typically available in the documentation or dataset used for training the YOLO model.

4.RESULT ANALYSIS

We use many Algorithms to perform Real-time object detection For Examples YOLO, SSD, CNN. In the following research we performed YOLO to predict the object detection. YOLO is known for its good speed and accuracy, it can process the whole image at once so we choosed this Algorithm to predict real time applications.

YOLO refers to You Look Only Once, for preprocessing the data it uses the necessary libraries PyTorch or TensorFow ,Darknet ,OpenCv ,Numpy ETC.

YOLO achieves high accuracy by using a convolutional neural network (CNN) to predict both the class and location of objects in an image. YOLO uses a fully convolutional network architecture, which allows for the efficient use of GPUs during training and inference.

YOLO uses anchor boxes, which allows the model to handle objects of different scales, thus allowing the model to detect objects of different sizes in the same image.

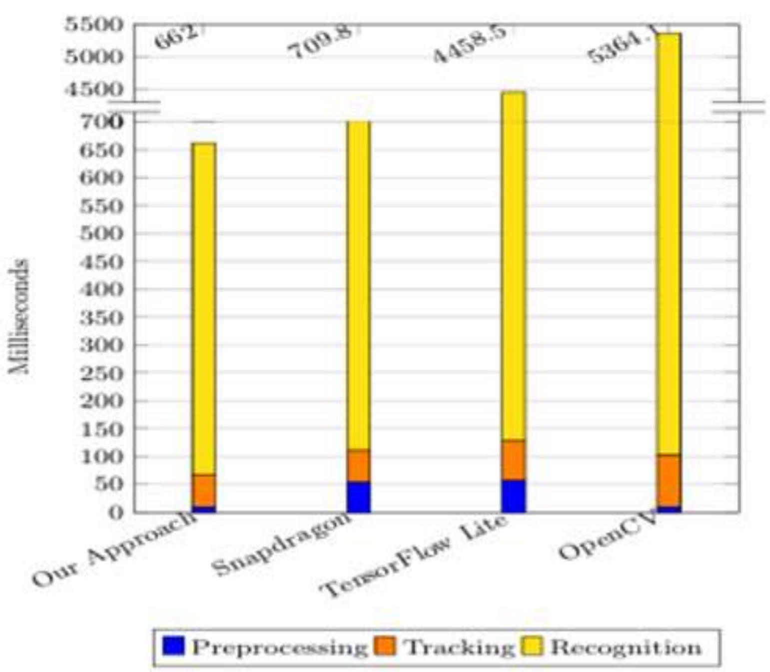

Fig 2: Object detection using different Systems

This graph compares the real-time performance of some different frameworks with regard to FPS over time. The key factor in object detection is FPS, since it determines how fast a system can process video frames for object detection and recognition. OpenCV is really exceptional in terms of processing speed, especially when real-time object detection is considered.
Snapdragon's performance varies mostly with the complexity of the task or the hardware conditions but has remained functional in most instances concerning real-time detection.
This would likely appear in fantastic detail with precisions or special detection models because the reason that leads to lower FPS but stable performance might possibly imply better handling in complex tasks of detection

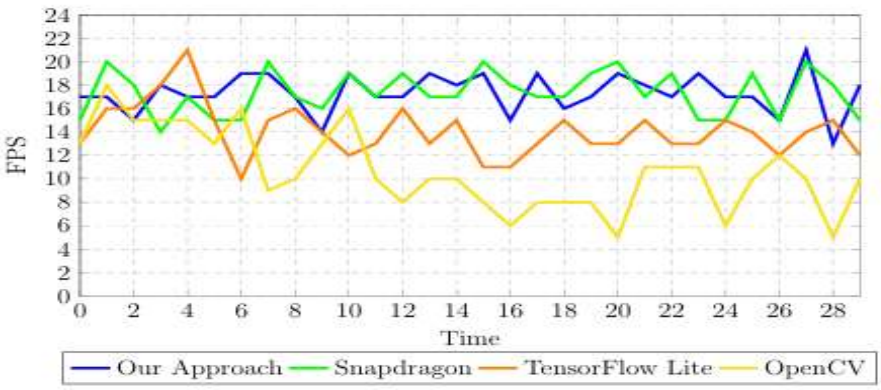

Fig 3: Different frames of object detection per second

This will give us an approximate form of the highest FPS stability; that, optimized in speed as well as constancy for real-time objects, shall be a decent performance.
Snapdragon can deliver high-speed constant real-time performance, so it's a great candidate to be leveraged either in an embedded application where reliance in the number of frames per second is needed for ensuring smooth object detection or in a mobile application.
There's another version, stable and working with this at the point of edge cases yet losing a little part of the FPS to efficiency and usage.
While OpenCV relishes its popularity in huge crowds to tap into so many computer vision applications, realizations of FPS are less stable; thus, it may be a sign of further optimizations on real-time object detection yet possible.

CONCLUSION

In this article we have explained YOLO object detection, one of the success stories in the modern deep learning approach to artificial intelligence. The YOLO algorithm classifies and locates multiple objects in one pass through an input image, making it fast enough to process real-time video.

YOLO is built upon a number of engineering innovations. It starts with a convolutional neural network, itself an important innovation, and adds elements to the ConvNet's output that specify the identity and location of multiple objects of various shapes and sizes.

Since YOLOv1 was described in the original 2015 paper, additional versions have been released that improve upon the original design, specialize the architecture for various applications, and take advantage of transfer learning and incompletely labeled training data.

In particular, YOLOv2 improved upon the original by taking advantage of higher resolution images, using batch normalization, and making greater use of anchor boxes. YOLOv3 added criteria for improved selection of bounding boxes, used variable levels of granularity to improve detection of small objects, and moved to Darknet-53, an improved Convent architecture.

YOLOv4 was the result of experimenting with and adopting a variety of modifications to make YOLO more efficient, such as batch normalization and MISH activation. YOLOv5 followed the YOLOv4 architecture, but offered four versions: small, medium, large, and extra-large, that offer trade-offs between training/run time and accuracy.

Of course, YOLO depends on training data – images annotated with object bounding boxes and classifications. To learn how to efficiently develop accurate training data sets, call merit and talk to an expert!

References:

[1] Patibandla, R. S. M. L., & Narayana, V. L. (2021). Computational intelligence approach for prediction of COVID-19 using particle swarm optimization. In *Advances in Computational Intelligence and Data Analytics* (Vol. 923, pp. 175–189). Springer. https://doi.org/10.1007/978-981-15-8534-0_9

[2] Markus Lochtefeld, Johannes Schoening, Michael Rosh, Antonio Kruger, LittleProjectedPlanet: An Augmented Reality Game for Camera Projector Phoneshttps://johannesschoening.de/me/

[3] Kim, S., & Lee, S. Estimating VR sickness caused by camera shake. IEEE Access, DOI: 10.1109/ACCESS.2020.3020791 https://www.dfki.de/~krueger/

[4] A Dice Game in Third-Person Augmented Reality, Richard Colvin, Ted Hung, David Jimison, Benjamin Johnson, Eben Myers, Tina Blaine Entertainment Technology Centre Carnegie Mellon University 700 Technology Drive Pittsburgh, PA I521 9 USA 5. MOLLA, E., AND LEPETIT, V. 2010. https://www.linkedin.com/in/mike-rosh-689b666

[5] LEITNER, J., HALLER, M., YUN, K., WOO, W., SUGIMOTO, M., AND INAMI, M. 2008. Inscrutable, a mixed reality tabletop game experience. In Proceedings of Advances in Computer Entertainment Technology (ACE), 9–16.

[6] https://www.researchgate.net/scientific-contributions/Markus-Lochtefeld-2215711928

[7] P. Beardsley, J. Van Baar, R. Rakkar, and C. Forlines. Interaction using a handheld projector. IEEE Computer Graphics and Applications, 25(1):39–43,200

[8] https://www.researchgate.net/profile/Alessandro-Dal-Corso

[9] A. D. Wilson. Play Anywhere: a compact interactive tabletop projection-vision system. In UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology, pages 83–92, New York, NY, USA, 2005. ACM.

[10] https://en.wikipedia.org/wiki/Gu%C3%B0mundur_and_Geirfinnur_ase

[11] V. N. Dao, K. Hosoi, and M. Sugimoto. A semi-automatic Realtime calibration technique for a handheld projector. In VRST '07: Proceedings of the 2007 ACM symposium on Virtual reality software and technology, pages 43–46, New York, NY, USA, 2007. ACM. https://scholar.google.co.in/scholar?q=yilmaz+et+al&hl=en&as_sdt=0&as_vis=1&oi=scholart

[12] William T. Freeman, David B. Anderson, Paul A. Beardsley, Chris N. Dodge, Michal Roth, Craig D. Weissman, and William S. Yearnings Computer Vision for Interactive Computer Graphics Article in IEEE Computer Graphics and Applications DOI: 10.1109/38.674971

[13] https://billf.mit.edu/about/bio

[14] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans.Pattern Anal. Mach. Intel. **39**(6), 1137–1149 (2017) https:// doi.org/ 10. 1109/ TPAMI. 2016. 25770 31

[15] B. Drayer, T. Brox, Object detection, tracking, and motion segmentation for object-level video segmentation, retrieved in 2016 from http://arxiv.org/abs/1608.03066

[16] Santhi Sri, K., Sandhya Krishna, P., Lakshman Narayana, V., & Khadherbhi, R. (2021). Traffic analysis using IoT for improving secured communication. In *Advances in Intelligent Systems and Computing* (Vol. 213, pp. 499–507). Springer. https://doi.org/10.1007/978-981-33-4443-3_48

[17] Narayana, V. L., & Bharathi, C. R. (2019). Multi-mode routing mechanism with cryptographic techniques and reduction of packet drop using 2ACK scheme MANETs. In *Advances in Intelligent Systems and Computing* (Vol. 104, pp. 649–658). Springer. https://doi.org/10.1007/978-981-13-1921-1_63

[18] Narayana, V. L., & Bharathi, C. R. (2018). Effective multi-mode routing mechanism with master-slave technique and reduction of packet droppings using 2-ACK scheme in MANETS. *Mathematical Modeling of Computer Systems*, 91(2), 73–76. https://doi.org/10.18280/mmc_a.910207

[19] Lakshman Narayana, V., Lakshmi Patibandla, R. S. M., Pavani, V., & Radhika, P. (2023). Optimized nature-inspired computing algorithms for lung disorder detection. In *Advances in Intelligent Systems and Computing* (Vol. 1066, pp. 103–118). Springer. https://doi.org/10.1007/978-981-19-6379-7_6

[20] Narayana, V. L., Sudheer, B. N., Maddumala, V. R., & Anusha, P. (2020). Fuzzy base artificial neural network model for text extraction from images. *Journal of Critical Reviews*, 7(6), 350–354. https://doi.org/10.31838/jcr.07.06.61

[21] Narayana, V. L., Bhargavi, S., Srilakshmi, D., Annapurna, V. S., & Akhila, D. M. (2024). Enhancing remote sensing object detection with a hybrid Densenet-LSTM model. In *Proceedings of the International Conference on Computer Science and Advanced Technology* (pp. 264–269). IEEE. https://doi.org/10.1109/IC2PCT60090.2024.10486394

[22] Narayana, V. L., Syamalatha, P., Vatsalya, P., Sricharitha, V., & Akhila, V. (2023). Multi-level node authorization using recurrent neural networks for secure health monitoring system. Proceedings of the IEEE International Conference on Smart Computing and Networking Applications (ICSNA), 1697–1705. https://doi.org/10.1109/ICSCNA58489.2023.10370543

[23] Gopi, A. P., Swathi, V., Harshitha, G. S., Swetha, B., & Alekhya, N. (2023). Prediction of paddy yield based on IoT data using GRU model in lowland coastal regions. In Proceedings of the 5th International Conference on Smart

Systems and Inventive Technology (ICSSIT 2023) (pp. 1747-1752). https://doi.org/10.1109/ICSSIT55814.2023.1006093
[24] Arepalli, P. G., Naik, K. J., & Amgoth, J. (2024). An IoT-based water quality classification framework for aqua-ponds through water and environmental variables using CGTFN model. International Journal of Environmental Research, 18(4), Article 73. https://doi.org/10.1007/s41742-024-00625-2
[25] Gopi, A. P., Babu, E. S., Raju, C. N., & Kumar, S. A. (2015). Designing an adversarial model against reactive and proactive routing protocols in MANETs: A comparative performance study. International Journal of Electrical and Computer Engineering, 5(5), 1111-1118. DOI: 10.11591/ijece.v5i5.pp1111-1118
[26] Sravanthi, G. L., Devi, M. V., Sandeep, K. S., Naresh, A., & Gopi, A. P. (2020). An efficient classifier using machine learning technique for individual action identification. International Journal of Advanced Computer Science and Applications, 11(6), 513-520. DOI: 10.14569/IJACSA.2020.011066
[27] Gopi, A. P., Durga Mani, P., Chandana, V. B., Sulthana, S. R., & Parameswari, P. P. K. (2024). Classification of fake news using enhanced capsule neural network. In 2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation, IATMSI 2024. DOI: 10.1109/IATMSI60426.2024.10502837

[28] Arepalli, P. G., & Khetavath, J. N. (2023). An IoT framework for quality analysis of aquatic water data using time-series convolutional neural network. Environmental Science and Pollution Research, 30(60), 125275-125294. https://doi.org/10.1007/s11356-023-27922-1
[29] Gopi, A. P., & Jairam Naik, K. (2021). A model for analysis of IoT based aquarium water quality data using CNN model. In 2021 International Conference on Decision Aid Sciences and Application, DASA 2021 (pp. 976-980). https://doi.org/10.1109/DASA53625.2021.9682251
[30] Roshini, P., Khajavali, S., Snigdha, M. L. S., Harsha, B. D., Srilakshmi, B., & Gopi, A. (2024). CNN design with AlexNet algorithm for diagnosis of diseases in cassava leaves. In Proceedings - 2024 International Conference on Expert Clouds and Applications, ICOECA 2024 (pp. 711-718). https://doi.org/10.1109/ICOECA62351.2024.00129
[31] Narayana, V. L., Gopi, A. P., Khadherbhi, S. R., & Pavani, V. (2020). Accurate identification and detection of outliers in networks using group random forest methodology. Journal of Critical Reviews, 7(6), 381-384. https://doi.org/10.31838/jcr.07.06.67
[32] Rao, B. T., Patibandla, R. S. M. L., Narayana, V. L., & Gopi, A. P. (2021). Medical data supervised learning ontologies for accurate data analysis. In Semantic Web for Effective Healthcare Systems (pp. 249-267). https://doi.org/10.1002/9781119764175.ch11
[33] Patibandla, R. S. M. L., Gopi, A. P., Narayana, V. L., & Rao, B. T. (2023). Decentralized smart healthcare systems using blockchain and AI. In Blockchain applications in healthcare: Innovations and practices (Vol. 1, pp. 139-154). DOI: 10.1002/9781394229512.ch8
[34] Lakshman Narayana, V., & Gopi, A. P. (2020). Enterotoxigenic Escherichia coli detection using the design of a biosensor. Journal of New Materials for Electrochemical Systems, 23(3), 164-166. DOI: 10.14447/jnmes.v23i3.a02
[35] Narayana, V. L., & Gopi, A. P. (2017). Visual cryptography for gray scale images with enhanced security mechanisms. Traitement du Signal, 34, 197-208. DOI: 10.3166/ts.34.197-208
[36] Arepalli, P. G., Narayana, V. L., Venkatesh, R., & Kumar, N. A. (2019). Certified node frequency in social network using parallel diffusion methods. Ingenierie des Systemes d'Information, 24(1), 113-117. https://doi.org/10.18280/isi.240117
[37] Peda Gopi, A., & Lakshman Narayana, V. (2017). Protected strength approach for image steganography. Traitement du Signal, 34(3-4), 175-181. https://doi.org/10.3166/TS.34.175-181
[38] Narayana, V. L., Gopi, A. P., Anveshini, D., & Lakshmi, G. V. V. (2020). Enhanced path finding process and reduction of packet droppings in mobile ad-hoc networks. International Journal of Wireless and Mobile Computing, 18(4), 391-397. https://doi.org/10.1504/IJWMC.2020.108539
[39] Challa, R., YAMPARALA, R., KANUMALLI, S. S., & KUMAR, K. S. (2020, November). Advanced patient's medication monitoring system with ardunio UNO and NODEMCU. In *2020 4th International conference on electronics, communication and aerospace technology (ICECA)* (pp. 942-945). IEEE.
[40] Kanumalli, S. S., Chinta, A., & Chandra Murty, P. S. R. (2019). Isolation of Wormhole Attackers in IOV Using WPWP Packet. *Revue d'Intelligence Artificielle*, *33*(1)
[41] Kanumalli, S. S., Ch, A., & Murty, P. S. R. C. (2018). Advances in Modelling and Analysis B. *Journal homepage: http://iieta. org/Journals/AMA/AMA_B*, *61*(1), 5-8.
[42] Kosaraju, Chaitanya, et al. "A model for analysis of diseases based on nutrition deficiency using random forest." 2022 7th International Conference on Communication and Electronics Systems (ICCES). IEEE, 2022.
[43] Chaitanya, Kosaraju, and Gnanasekaran Dhanabalan. "Secure Route Detection with Multi Level Trust Evaluation Model Using Replicated Auditor Node for Extended Packet Delivery Rate in WSN." Revue d'Intelligence Artificielle 37.4 (2023).
[44] Chaitanya, Kosaraju, et al. "Risk Stratification for Stroke Using Attention Transformer Model." 2024 2nd International Conference on Disruptive Technologies (ICDT). IEEE, 2024.