International Journal of Computer Sciences and Engineering
Vol.**12**, Issue.**4**, pp.**75-80**, April **2024**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

Research Article

Performance Evaluation on Real-time object detection using DL techniques

B. Sai Jyothi[1], Chavali Saathvika Durga Abhinaya[2*], Bellamkonda Lahari[3], Chinta Devika Priya[4], Devarapalli Anjali[5], Bathula Sri Navya[6]

[1,2,3,4,5,6]Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, AP, India

**Corresponding Author: saathvichavali@gmail.com*

Received: 05/Mar/2024; **Accepted**: 06/Apr/2024; **Published**: 30/Apr/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i4.7580

Abstract: Objects are located by drawing a bounding box around the detected object. One of computer vision's specialties is object detection, which finds things in an image or video. Techniques for object detection are the foundation of the area of artificial intelligence. Typically, object detection uses deep learning and machine learning to yield accurate and significant findings. It is essentially made up of localization and classification. The state-of-the-art techniques utilized for real-time object detection have advanced recently. This study paper compares state-of-the-art techniques, such as faster region convolutional neural networks (Faster R-CNN) and you only look once V8 (YOLOV8). These algorithms are deep neural network representations, or neural networks with numerous hidden layers. Although each of these algorithms is notable for its own distinctiveness, they are compared to see which is superior. This study focuses on determining which of these algorithms is more practical to employ despite sharing a common core, namely CNNs.

Keywords: YOLOV8(You only look once), Faster region convolutional neural network (faster R-CNN), object detection, deep learning, deep neural networks, and convolutional neural networks.

1. Introduction

Since humans have such a keen sense of vision, they can quickly identify and recognize items in their environment, regardless of their position or colour. However, object detection on computers is more difficult and involves more processing. The study of how computers comprehend digital photos or movies at a high level is known as computer vision. The components of computer vision are image categorization, picture captioning, picture identification, etc.

Essentially, object detection serves as the cornerstone of artificial intelligence. The most widely used deep learning technology is the convolutional neural network (CNN), which uses many convolutional layers and convolutional computation to improve detection accuracy and speed. All algorithms for object detection rely on convolutional neural networks.

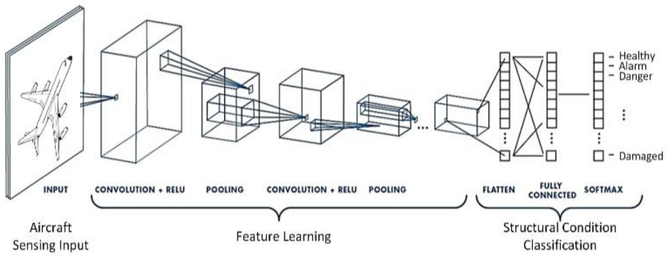


Figure 1: Convolutional Neural Network Architecture.

A deep convolutional neural network, such as CNN, is one of the artificial neural networks that combines convolutional layers with other types of layers, including nonlinear, pooling, and fully connected layers. It trains its convolutional filters through the use of backpropagation.

In the present research, the more accurate and efficient deep learning algorithms are YOLOV8 (you only look once) and Faster R-CNN (Regional Convolutional Neural Networks). In object detection.

There are two types of detection algorithms: the first is a one-step method called You Only Look Once (YOLOV8) and the second is a two-step technique called Faster Region Based Convolutional Neural Network (Faster R-CNN). The two processes are object categorization, where items are categorized according to the colours they have and object localization, in which drawing a bounding box around an object that has been recognized helps locate it.

2. YOLO: A Brief History

The popular object identification and picture segmentation approach known as YOLO (You Only Look Once) was created at the University of Washington by Joseph Redmon and Ali Farhadi. Yolo was introduced in 2015 and immediately became well-known for its rapid speed and accuracy.

The 2016 release of YOLOv2 enhanced the original model by adding dimension clusters, anchor boxes, and batch normalization.

The 2018 release of YOLOv3 improved the model's performance even more by employing spatial pyramid pooling, numerous anchors, and an even more effective backbone network.

When YOLOv4 was launched in 2020, it included additional features such a new loss function, an anchor-free detection head, and mosaic data augmentation.

The model's performance was further enhanced by YOLOv5, which also included new capabilities including integrated experiment tracking, automatic export to widely used export formats, and hyperparameter optimization.

Many of Meituan's autonomous delivery robots utilize YOLOv6, which was made available as open source software in 2022.

Additional tasks, like posture estimation on the COCO keypoints dataset, were added to YOLOv7.

2.1 YOLOV8

Ultralytics has released YOLOV8, the most recent version of YOLO. As a state-of-the-art, cutting-edge (SOTA) model, YOLOv8 expands upon the features and enhancements of earlier iterations to offer improved efficiency, performance, and flexibility. A complete YOLOv8 support several AI vision tasks, such as tracking, position estimation, segmentation, detection, and category.

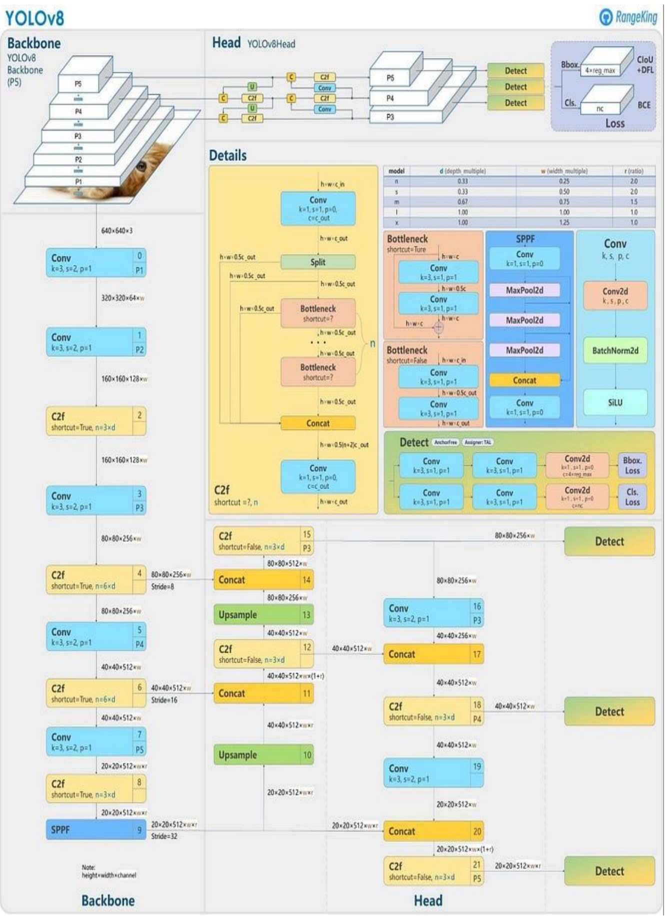


Figure 2: YOLOV8 Architecture

Since YOLOv8 is so adaptable, users can take advantage of its features in a variety of contexts and fields. Here is the most recent iteration of the well-liked real time object identification and image segmentation model, Ultralytics YOLOv8. YOLOv8 offers unmatched speed and accuracy thanks to state-of-the-art developments in deep learning and computer vision. It may be easily adapted to many hardware platforms, ranging from edge devices to cloud APIs, thanks to its simplified design, which makes it suited for a wide range of applications.

2.2 Faster RCNN

One of the most popular and widely used R-CNN variants is faster R-CNN. It suggests areas based on a particular set of search methods that need CPU computation and take a few seconds (or more) for each picture. The Faster R-CNN generates region proposals using Region Proposal Networks, or RPNs, which reduces the picture generation time from seconds to milliseconds.

-In Bounding boxes, or rectangular boxes that ring an object and indicate its position, class (such as person or automobile), and confidence (i.e., how likely it is to be at that location), are generated by Faster R-CNN using RPN.

-CNN is usually used to generate these objects' features during this stage. Instead of using the original image, which is then fed into ROI pooling to modify the image size required for object recognition, the final feature picture is what is used for region proposal.

-The ROI pooling layer's output size is (N, 7, 7, 512), whereas the region proposal technique yielded N recommendations overall. The ROI pooling outputs are fed into the sibling classification and regression branches after they have gone through two fully linked layers. A layer of classification is present to determine the class to which the object belongs.

-The bounding box coordinates are then adjusted even further, removing all possibility of inaccuracy, by use of a regression layer. To manage objects with different sizes and aspect ratios, anchors are added in RPN. An anchor is a sliding point of the convolutional maps that is situated in the center of each spatial window.

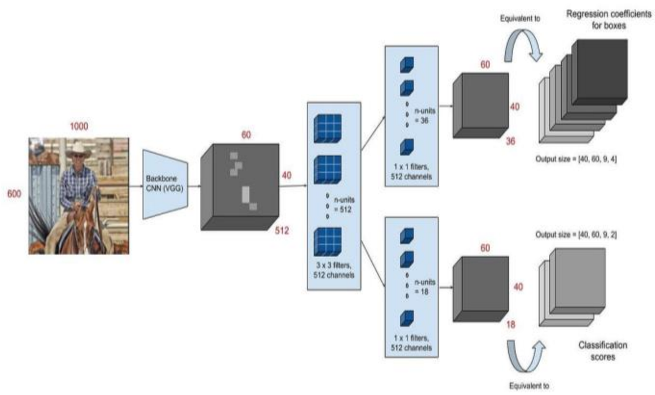


Figure 3: Faster R-CNN architecture

3. Theory/Calculation

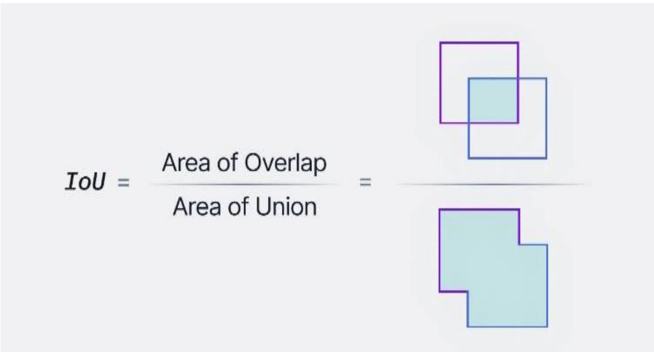
We require common quantitative criteria to assess and contrast the predicted performance of various object detection methods. The most widely used evaluation measures include recall, MAP, Average Precision (AP) metrics, Intersection over Union (IoU), and so on. To compare the algorithms in this case, we employed IOU metrics.

A common metric used to determine localization faults and assess localization accuracy in object identification models is intersection over union. We initially take the intersection area between the two matching bounding boxes for the same object in order to calculate the IoU between the predicted and the ground truth bounding boxes.

Next, we determine the overall area that the two bounding boxes—also referred to as the "Union"—cover as well as the area where they overlap, or the "Intersection." We can determine how near the prediction bounding box is to the original bounding box by dividing the intersection by the Union, which yields the ratio of the overlap to the entire area.

3.1 Intersection over Union (IoU):
Intersection over Union is a popular metric to measure localization accuracy and calculate localization errors in object detection models. To calculate the IoU between the predicted and the ground truth bounding boxes, we first take the intersecting area between the two corresponding bounding boxes for the same object.

Following this, we calculate the total area covered by the two bounding boxes also known as the “Union” and the area of overlap between them called the “Intersection”. The intersection divided by the Union gives us the ratio of the overlap to the total area, providing a good estimate of how close the prediction bounding box is to the original bounding box.



3.2 Bounding Box Coordinates:
A rectangular box that contains an object found in a picture is called a bounding box. The following are the crucial coordinates connected to a bounding box:

1. Coordinates of X and Y (Box Center):
(x, y) is a representation of the coordinates of the bounding box's center inside the picture. These figures are in relation to the image's dimensions.

2. Width and Height (Box Dimensions):
Height and breadth show the bounding box's measurements. They ascertain the box's length along the x- and y-axes, respectively.

3. Top-Left and Bottom-Right Coordinates:
The bounding box's top-left and bottom-right corners' coordinates are represented by the variables (x_min, y_min) and (x_max, y_max). The center, width, and height are used to calculate these coordinates.

3.3 Anchor Box Coordinates:
When it comes to YOLO and anchor boxes, each one is distinguished by

1. Having its own unique set of coordinates: Additionally Elevation of the Anchor Box: The size of the anchor box are indicated by the variables anchor_width and anchor_height. These numbers are preset and are used as models.
2. Aspect Ratio while receiving instruction. Anchor Box: LSTM use a gateway mechanism to control the quantity of previous recordings preserved by each LSTM component, recall the current input, preserve important characteristics, and reject insignificant ones. This solves the vanishing problem in recurrent neural networks. Figure 2 shows the component model of the LSTM. anchor box coordinates relate to the specifications that specify the position and dimensions of anchor boxes within of an image. Typically, the values for these coordinates are expressed in relation to the grid cell where the anchor box is located. Anchor boxes are commonly specified in YOLO and related object identification algorithms as follows:

Center coordinates (x, y): These show where the anchor box's center is in relation to the grid cell. Normalization of these values typically takes place within the interval [0, 1], where the top-left corner of the grid cell is represented by (0, 0) and the bottom-right corner by (1, 1).

The anchor box's dimensions are determined by its width (w) and height (h). Moreover, they are commonly normalized to the interval [0, 1].

4. Comaprision between YOLOV8 and Faster R-CNN

YOLOV8 (You Only Look Once) and Faster R CNN are two well-liked object identification techniques in computer vision, each with unique benefits and features.

YOLOV8 is well renowned for its efficacy in realtime application, despite the fact that it anticipates bounding boxes and class probabilities with a single neural network pass and treats object recognition as a single regression problem-solving In contrast, Faster R-CNN has a two stage detection pipeline that divides the duties of object categorization and region proposal creation.

Faster R-CNN tends to generate fewer IOU measurements than YOLOV8, which usually results in a higher detection method..

Figure 4: Comparison of two algorithms

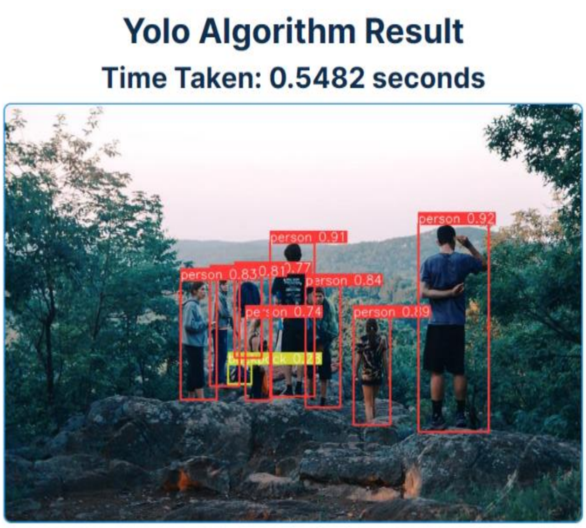
5. Results and Discussion

When it comes to object detection using IOU metrics, YOLOv8 usually outperforms Faster R-CNN, exhibiting higher Intersection over Union (IOU) scores, which suggest more precise bounding box predictions and improved object localization. Yolo is the recommended option for real-time applications due to its efficiency and efficacy.

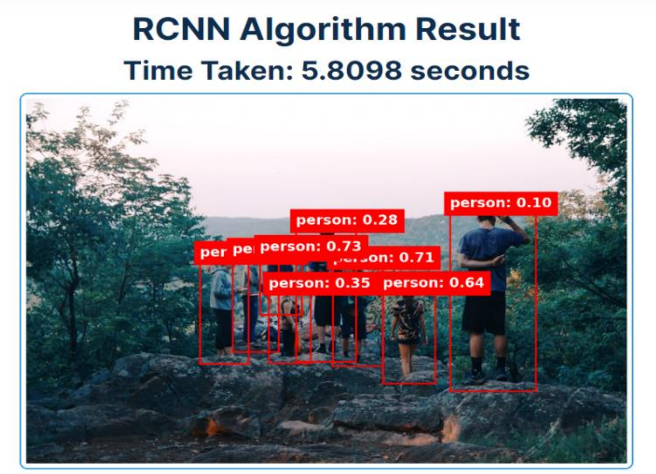
The object detection and IOU values utilizing Yolov8 and faster RCNN are displayed in the images below

Testcase1:

- **Detection with YOLOV8**

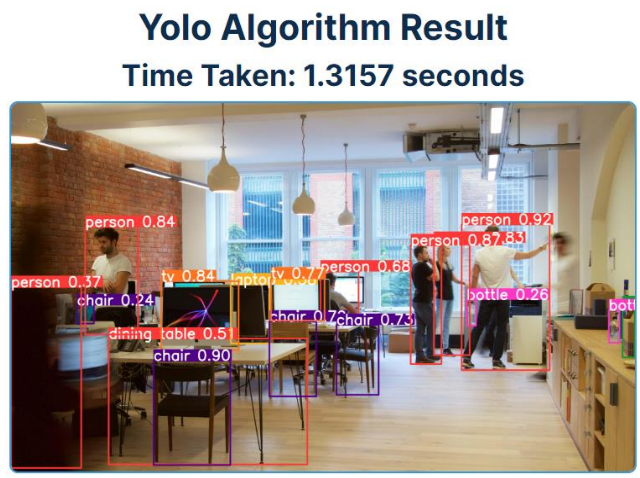


- **Detection with Faster RCNN**

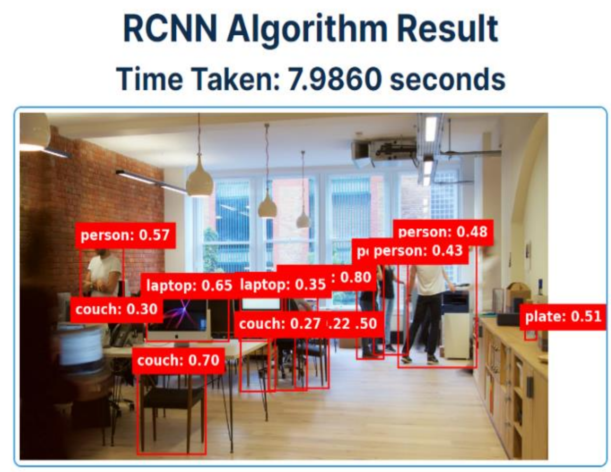


Testcase2:

- **Detection with YOLOV8**



- **Detection with Faster RCNN**



Ultimately, both YOLOv8 and Faster R-CNN represent powerful tools in the field of object detection. But when compared to Faster RCNN, YOLOV8 detects objects with quick time and gives the high intersection over union(IoU) metrics value.

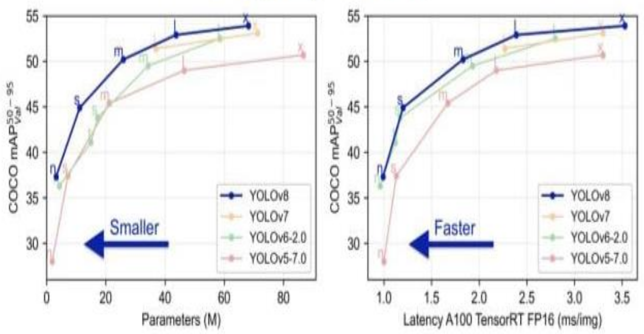
Comparison Table:

6.2.7 PERFORMANCE EVALUATION METRICS

LABELS	IOU METRICS	
	YOLOV8	FASTER RCNN
Airplane	0.94	0.58
Boat	0.91	0.45
Laptop	0.84	0.47
Chair	0.96	0.72
Car	0.84	0.60
Truck	0.89	0.53
Traffic lights	0.87	0.21
Bed	0.69	0.54
Potted Plant	0.92	0.53
Keyboard	0.89	0.67
Mouse	0.96	0.70
Refrigerator	0.89	0.48
Elephant	0.92	0.57
Bear	0.80	0.66
Giraffe	0.88	0.58
Apple	0.91	0.59

Graph:

Graph Representation



6. Conclusion

Anchor boxes play a crucial role in improving the robustness and accuracy of object detection models like YOLOV8. They offer an organized method for the a model to pick up on and adjust to the various object sizes and aspect ratios in a given dataset, resulting in higher accuracy and dependability in object location. YOLOV8 has more advanced uses than Faster R-CNN.As YOLOV8 offers end-toend training, it turns out to be a cleaner and more effective method for doing object detection. While both algorithms have an achievable level of accuracy, there are instances where YOLOV8 performs faster, more accurately, and more efficiently than Faster R-CNN. Yolov8 is more suited for real-time object detection in images and videos since it can execute single-shot algorithms. It is easy to build and can be trained on entire images directly. When it comes to object generalization, YOLOV8 outperforms Faster R-CNN, making it a more reliable, quick, and strong method. This method stands out and is highly recommended because of these tremendous advantages.

References

[1] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” arXiv, Jan. 06, 2016. Accessed: Jun. 17, **2023**. [Online]. Available: http://arxiv.org/abs/1506.01497.

[2] S. J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement.” arXiv, apr. 08, 2018. Accessed: **Sep. 25, 2022**. [Online]. Available: http://arxiv.org/abs/1804.0276.

[3] A.M.A.ghani Abdulghani and G.G. Menekse Dalveren, “Moving Object Detection in Video with Algorithms YOLO and Faster R-CNN in Different Conditions,” European Journal of Science and Technology, **Jan. 2022**, DOI: 10.31590/ejosat.1013049.

[4] H. Jiang and E. Learned-Miller, Face Detection with the Faster R-CNN.” arXiv, Jun. 10, 2016. Accessed: **Sep. 25, 2022**. [Online].Available: http://arxiv.org/abs/1606.03473.

[5] Chandana, R. K., & Ramachandra, A. C. Real time object detection system with YOLO and CNN 740 models: A review. arXiv preprint arXiv:2208.00773, **2022.**

[6] JiayiFan; JangHyeon, Lee; InSuJung; YongKeunLee, “Improvement of Object Detection Based on Faster R-CNN and YOLO”, International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC) **June,** pp.**27-30 2021.** DOI: 10.1109/ITC-CSCC52171.2021.9501480.

[7] J. Kim, J.-Y. Sung, and S. Park, “Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition,” in 2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), Seoul, Korea (South): IEEE**, Nov.** pp.**1–4**, **2020.** DOI: 10.1109/ICCE-Asia49877.2020.9277040.

[8] F. Miao, Y. Tian, and L. Jin, “Vehicle Direction Detection Based on YOLOv3,” in 2019 11th International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC), Hangzhou, China: IEEE, **Aug.**, pp.**268–271, 2019**. DOI: 10.1109/IHMSC.2019.10157.

[9] Rohan, A., Rabah, M., & Kim, S. H. Convolutional neural network-based real-time object detection and tracking for parrot AR drone 2. IEEE access, 7, pp.**69575- 69584, 2019.**

[10] Younis, A., Shixin, L., Jn, S., & Hai, Z. (January). Real-time object detection using pre- trained deep learning models MobileNet-SSD. In Proceedings of 2020 6th International Conference on Computing and Data Engineering, pp.**44-48, 2020.**

[11] Nguyen, N. D., Do, T., Ngo, T. D., & Le, D. D. An evaluation of deep learning methods for small object detection. Journal of electrical and computer engineering, 2020, pp.**1-18, 2020.**

[12] Hossain, S., & Lee, D. J. Deep learning-based real- time multiple-object detection and tracking from aerial imagery via a flying robot with GPU-based embedded devices. Sensors, Vol.**19**, Issue.**15**, pp.**33-71, 2019.**

[13] Pal, S. K., Pramanik, A., Maiti, J., & Mitra, P. Deep learning in multi-object detection and tracking: state of the art. Applied Intelligence, 51, pp.**6400-6429, 2021.**

[14] J. Du, Understanding of Object Detection Based on CNN Family and YOLO,” J. Phys.: Conf. Ser., **Apr.,** Vol.1**0**04, pp.**12-29**, **2018.** DOI: 10.1088/1742-6596/1004/1/012029.

[15] S. Ren, K. He, R. Girshick, and J. Sun, Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Trans. Pattern Anal. Mach. Intell., **Jun.,** Vol.**39**, No.**6**, pp.**1137–1149, 2017**, DOI: 10.1109/TPAMI.2016.2577031.

AUTHORS PROFILE

B. Sai Jyothi working as Professor Department of Information Technology, Vasireddy Venkatadri Institute of Technology,Guntur.

Chavali Saathvika Durga Abhinaya B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur.

Bellamkonda Lahari B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur.

Chinta Devika Priya B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur.

Devarapalli Anjali B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur.

Bathula Sri Navya B. Tech Students Department of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur.

