International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-5, June 2019

Fine-Tuning Convolutional Neural Network Models for improvement of Object Detection Accuracy

Chamarty Anusha, P. S. Avadhani

Abstract: ***Deep Learning is getting a lot of attention these days as these algorithms are out-performing humans in Object Detection. It has entered into several diverse areas like Automobile Industry for making self - Driving cars, Automation industry in Robot development, Medical field for disease identification and for security enhancement in defense sector etc. Deep Learning has become the solution for a variety of problems like Natural language processing, Speech Recognition and Visual Recognition [1,2] etc. Advancement in Deep Learning is especially due to the improvement of Convolution Neural Network Architectures and implementation of new algorithms. Deep Learning is a class of Artificial Neural Networks that learn features from the weights obtained from neurons. In this paper a dataset of animal is taken and it is passed through different Deep Learning Neural Network models like Visual Geometry Group-16 (VGG-16), Visual Geometry Group-19 (VGG-19), Inception, Xception and their accuracy of prediction is compared. To increase the accuracy of prediction further, a method named Fine-Tuning [3,4,5] is applied. Fine-Tuning Neural Network model means varying various hyper-parameters like 'Number of Top Layers to be Trained', 'Learning rate', 'Number of Epochs', Drop-out rate, Optimization algorithm, Activation Function, Training Batch Size, Validation Batch Size, Testing Batch Size, Training dataset Size, Validation dataset Size, Testing Dataset Size etc. and their accuracies of prediction before and after Fine-Tuning are analyzed. There is a significant improvement observed in accuracy of prediction of considered animal dataset after Fine-Tuning of the various Neural Network Models. Based on the experimental results obtained, it can be concluded that the Object Detection accuracy can be enhanced and the detection error rate can be reduced to a greater extent by fine-tuning method.***

Keywords: VGG16, VGG19, Inception, Xception.

I. INTRODUCTION

Convolutional Neural Networks (CNNs) [1,6,7] are made up of neurons, which learn weights from training and have a constant bias value. Each neuron receives several inputs from its previous layer and it performs the weighted sum of all those inputs, which are then passed through an activation function. The output of this function is further passed through the fully connected layer of a Neural Network, and this layer responds by specifying the classification of the image. CNNs came into existence primarily after the improvement of computational power of computers.

Manuscript published on 30 June 2019.
***** Correspondence Author (s)
Chamarty Anusha, Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam (AP), India.
P S Avadhani, Computer Science & Systems Engineering, Andhra University College of Engineering, Andhra University, Visakhapatnam (AP), India.

Also, the CNNs usage has been drastically improved with the vast availability of datasets for training and testing.
CNNs are made up of four layers.

- Convolution Layer- for multiplication of inputs with weights of the network.
- Rectified Linear Unit (ReLU) – Activation Function for addition of Non-Linearity.
- Pooling Layer - Max/Average Pooling for reducing the spatial size of representation.
- Fully Connected Layer – Softmax Layer for the purpose of prediction or classification of images.

Above mentioned CNNs are implemented in python language with Keras[8] package, which is a high level library and TensorFlow [8] package, which acts as back end. There are various Convolutional Neural Network models available in literature. A brief description of some of them is given below along with their Top-1 & Top-5 accuracies when they are executed with ImageNet Dataset. ImageNet [12], is a standard dataset having 15 million labeled high-resolution images with around 22,000 categories. In total, there are around 1.2 million training images, 50,000 validation images and 150,000 testing images.

A. Visual Geometry Group-16:

Visual Geometry Group (VGG-16) is a Convolutional Neural Network proposed by K. Simonyan and A. Zisserman in their paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" [9] and this model is tested on Imagenet Dataset. VGG-16 architecture has sixteen Convolution layers, five Max Pooling layers, three Fully Connected layers and a Softmax layer as a final layer for image classification. Rectified linear units are applied to each of the hidden layers. Input to the Convolution layer1 is 224*224 RGB image. The filters used in this model are 3*3 pixels with a stride of 1*1 pixel size. Max Pooling is performed over 2*2 pixels with a stride of 2. Fully connected layers are present at the end of the network. In total, three fully connected layers are present and out of which two layers have 4096 channels each and the third layer has 1000 channels and Softmax layer is present at the end. VGG-16 obtains an error rate of 8.8% and the network's accuracy improved by adding 16 Convolutional layers. The number of parameters of this model is 138 million.
The accuracies obtained are

- Top-1 Accuracy: 70.5%
- Top-5 Accuracy : 90.0%

B. Visual Geometry Group-19:

Visual Geometry Group-19 (VGG-19)[9] is an improved version of VGG-16. It was introduced by Visual Geometry Group of the University of Oxford. The input to Convolution layer1 of VGG-19 is a 224*224 RGB image. VGG-19 has 19 layers of deep network and can classify images of 1000 classes. This network contains 3*3 Convolutional layers stacked one over the other in increasing depth. The number of parameters of this model is 144million.

The accuracies obtained by this model are

- Top-1 Accuracy: 75.2%
- Top-5 Accuracy: 92.5%

The main drawbacks with VGGNet (VGG-16 & VGG-19) are

- Takes lot of time for training.
- The network architecture weights are very huge in size i.e., 533MB space with respect to disk space.

C. Inception

Inception was proposed by a team of scientists - Christian Szegedy**,** Wei Liu, Yangqing Jia, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke and Andrew Rabinovich in their paper titled "Going Deeper with Convolutions" [10] in 2014. The input to the convolution layer of Inception model is 299*299 RGB image. The Inception architecture was originally called as GoogleNet and it was later renamed as Inception-V1, which is 27 layers deep. This version was refined by the addition of batch normalization technique and renamed as Inception-V2. Again this version was further revised by the addition of factorization and more layers. That is the number of layers had been increased from 27 to 48 and this version is called Inception-V3. Inception-V1 model has approximately 7 million parameters which is very less when compared with the VGGNet models described above.

Accuracies obtained by this model on Imagenet Dataset are

- Top-1 Accuracy : 78.8%
- Top-5 Accuracy : 94.4%

D. Xception

Xception is extreme version of Inception. Xception was proposed by Francois Chollet in his paper titled "Deep Learning with Depth wise Separable Convolutions" [11]. It has a modified depth wise separable convolution and it is a better version than Inception-V3. Xception slightly outperforms Inception-V3 on Imagenet Dataset and it performs well on very large datasets like 350 million images of 17000 classes. Improved accuracy of Xception model is because of the efficient usage of model parameters. Xception model is a linear stack of 36 depth-wise separable Convolutional layers structured into 14 modules and all of them have linear residual connections around them. The input to the convolution layer of Xception model is 299*299 RGB image and it has the same number of parameters as Inception model. The weights of this model are of 91MB size and it is the smallest weighted model.

Accuracies obtained by this model are

- Top-1 Accuracy : 79%
- Top-5 Accuracy : 94.5%

II. METHODOLOGY

In this paper, an animal Image Dataset is taken for analysis purpose and it is passed through the above referred models and their prediction accuracy is analyzed. Initially a dataset of 151 images is passed through each Neural Network Model and their prediction accuracy along with the number of wrong predictions are analyzed and tabulated in Table-1.

As seen from Table-1, the accuracy of prediction is not satisfactory and the number of wrong predictions is also more. In order to increase the accuracy of prediction, Fine-Tuning [2,4,5] is applied on the taken neural network models and their prediction accuracy is analyzed.

A. Fine-Tuning

Fine-tuning [3,4,5] is a process of taking the weights of pre-trained neural networks like VGG-Net, Inception, Xception for training a new dataset for increasing the accuracy of Object Detection.

Fine-tuning is generally applied to the top layers of the network rather than the bottom layers as these layers contain more common features like edges, color blobs, lines etc. Top layers of the networks contain more specific features, which help in accurate Object Detection.

There are numerous advantages of Fine-Tuning process and some of them are

- Speeding up the training process, as there is no need to train the new dataset from the scratch.
- Training with a small dataset also shows improved accuracy.

III. EXPERIMENTAL RESULTS

This paper is an analysis of the above referred Convolutional Neural Network Models. In this paper a Dataset of 1740 animal images were taken for training purpose and 151 images were taken for Validation purpose. All these images are of 19 different animal sets.

There are various abbreviations used in the analysis and the same were as mentioned below.

NNM - Neural Network Model
PA% - % of Prediction Accuracy
VA% - % of Validation Accuracy
TA% - % of Training Accuracy
WP - Number of Wrong Predictions per 151 validation Images

A. Prediction Accuracy Prior To Fine-Tuning:

As discussed above, a dataset of 151 validation images are passed through each model and their accuracy in detecting the animals is tabulated as shown in Table-1.

Table 1 - NNM accuracy before Fine-Tuning

NNM	PA%	WP
VGG16	70.86	44
VGG19	72.18	42
InceptionV3	80.13	30
Xception	82.78	26

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

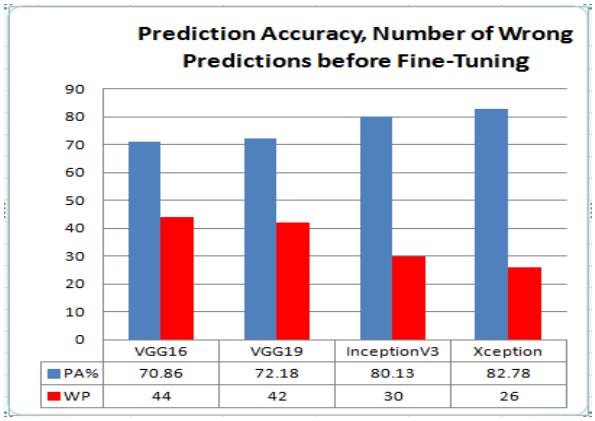

Figure 1- Prediction Accuracy and Number of Wrong predictions before Fine-Tuning.

It was identified that the VGG-16 model is giving more number of wrong predictions and less prediction accuracy while, the Xception model is giving more accurate prediction and the number of wrong predictions were also less.

But, the highest prediction is only 82.78%, which is not an acceptable value for accurate object detection. So, in order to improve the accuracy of prediction to a satisfactory limit, Fine-tuning is applied on the above Neural Network Models.

B. Experimental Results after Fine-tuning:

The following steps are followed for Fine-Tuning of NNMs:

1. Various executions were performed for identification of better values to each hyper-parameter and the same were freezed in all considered network models.
   - Learning rate: '$1e^{-4}$'
   - Training batch size: 100
   - Validation batch size: 10
   - Drop-out rate: 0.5
   - Activation Function: Relu[12]
   - Training Layers: Top-4 Layers
2. "Number of Epochs" is a hyper parameter, which is kept variable for the analysis.
3. The outputs for various epoch rates like 10, 20 & 50 are considered for analyzing the NNMs.

➢ Epoch rate – 10:

The outputs are analyzed and tabulated in Table-2.

Table 2 - NNM accuracy after Fine-Tuning for 10-epochs

NNM	TA%	VA%	WP
VGG16	87.97	91.39	13
VGG19	82.44	94.04	9
InceptionV3	92.52	95.36	7
Xception	97.14	100	0

➢ Epoch rate – 20:

To increase the prediction accuracy, "Number of Epochs" was increased to '20'. The outputs received are tabulated in Table-3.

Table 3 - NNM accuracy after Fine-Tuning for 20-epochs

NNM	TA%	VA%	WP
VGG16	94.44	98.68	2
VGG19	92.83	96.69	5
InceptionV3	94.60	98.01	3
Xception	98.23	99.34	1

➢ Epoch rate – 50:

In this case the numbers of epochs were further increased and set to '50' to achieve accuracy close to 100% and got satisfactory results. The outputs received are tabulated in Table-4.

Table 4 - NNM accuracy after Fine-Tuning for 50-epochs

NNM	TA%	VA%	WP
VGG16	100	100	0
VGG19	98.22	100	0
InceptionV3	92.27	99.34	1
Xception	99.67	100	0

C. Output Screenshots

The following are the screenshots of wrong predictions (WP) outputted by the NNMs after Fine-Tuning.

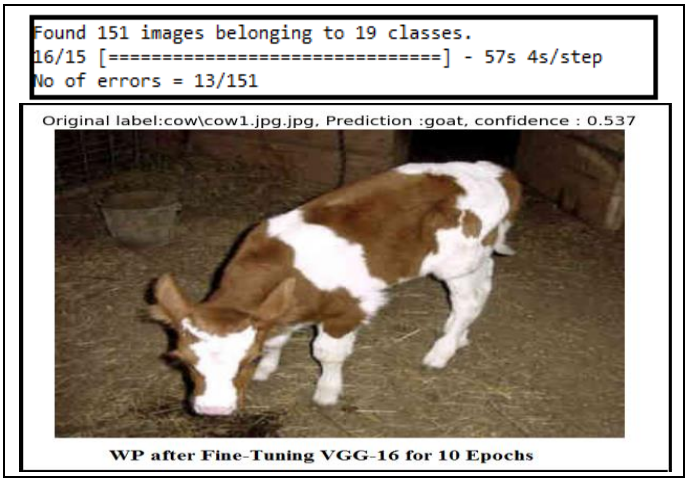

Figure 2: Wrong prediction by VGG-16

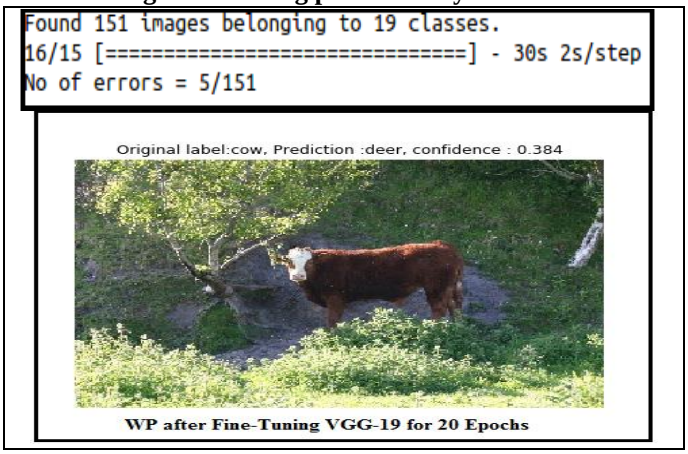

Figure 3: Wrong prediction by VGG-19

Figure 4: Wrong prediction by InceptionV3

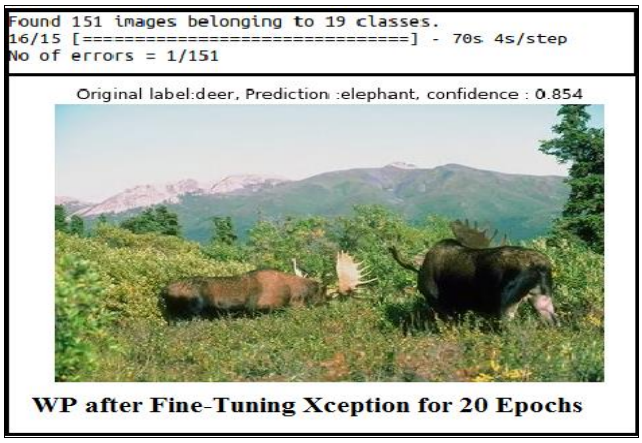

Figure 5: Wrong prediction by Xception

IV. DISCUSSIONS

From the results obtained, various comparisons were made for analyzing the training accuracy, validation accuracy and the number of wrong predictions with respect to the number of epochs. The Figure-6 shows the comparison of training accuracy by varying the number of epochs as 10, 20 and 50. It was identified that the training accuracy improved in all the Neural Network Models as the number of epochs increased.

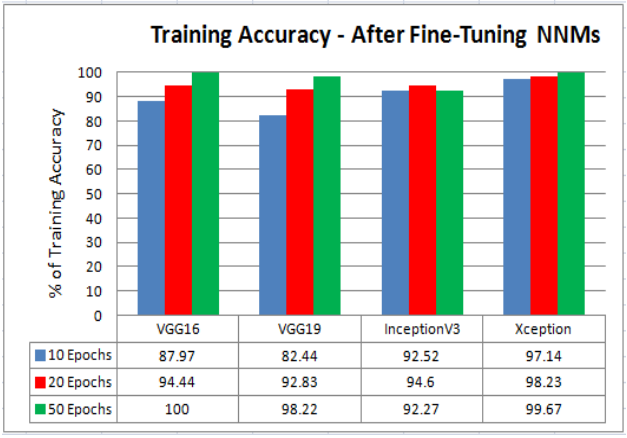

Figure 6: Training Accuracy after Fine-Tuning NNMs

The Figure-7 shows the comparison of validation accuracy with respect to the number of epochs. The validation accuracy improved in almost all the NNMs as the number of epochs increased and it reached 100% for majority of NNMs as the epoch rate approached 50.

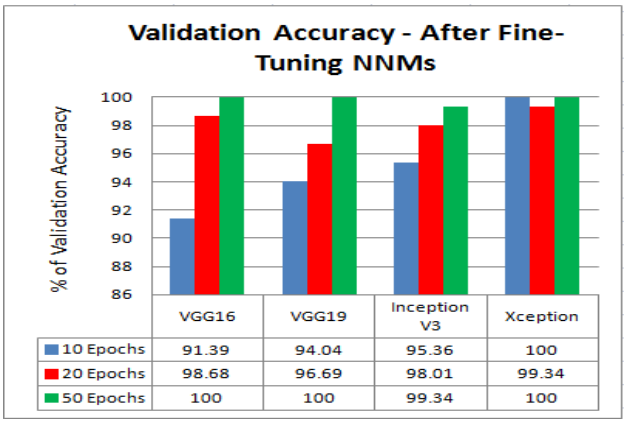

Figure 7: Validation Accuracy after Fine-Tuning NNMs

The Figure-8 shows the comparison of number of wrong predictions as epoch rate is varying. The wrong predictions were reduced in all the NNMs as the epoch rate increases. The number of wrong predictions reached zero as epoch rate is increased.

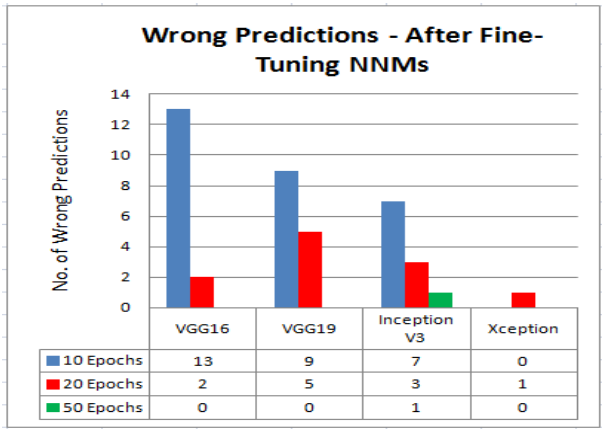

Figure 8: Wrong Predictions after Fine-Tuning NNMs

By analyzing the results, it is clear that after fine-tuning, the number of wrong predictions has been drastically reduced and even became zero for some NNMs. The same is clearly shown in Figure-9.

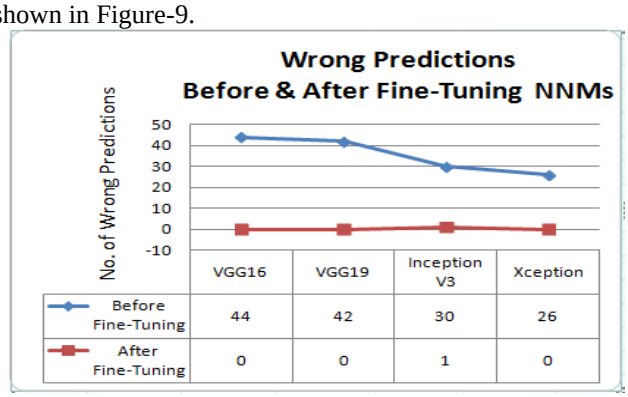

Figure 9: Wrong Predictions Before & After Fine-Tuning NNMs

V. CONCLUSION

Pre-trained Deep Learning Convolutional Neural Network models available in literature are designed mainly for the purpose of Image Classification, Object Detection and Object Recognition.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Each model has a specific level of accuracy in achieving this task. To improve the accuracy of the existing models, based on the data from the obtained results, it can be concluded that Fine-Tuning is the best solution as there is no need to train the Neural Network from scratch, which needs a lot a computational power and several days of training. In this paper Fine-Tuning is applied by changing only the hyper-parameter "Number of epochs" and keeping the rest of the parameters unchanged. It is clear from the results that by increasing the Number of epochs object detection accuracy has been considerably increased and reached 100% in majority of the NNMs.

VI. FUTURE SCOPE

In this paper, Fine-Tuning of a NNM is performed by changing only the hyper -parameter "Number of epochs". To achieve better accuracy, the affect of variation of other hyper parameters also needs to be further analyzed.

VII. ACKNOWLEDGMENT

This work is done under the scholarship of Visvesvaraya PhD Scheme for Electronics and IT, Government of India.

REFERENCES

1. Yann LeCun, Yoshua Bengio, Geoffrey Hinton," Deep Learning" in Nature, vol. 521, no. 7553, May 2015, pp. 436–444.
2. Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, GangWang, Jianfei Cai, Tsuhan Chen, "Recent Advances in Convolutional Neural Networks", in Pattern Recognition, vol. 77, May 2018, pp. 354-377
3. Ross Girshick, Jeff Donahue, Trevor Darrell, Jotendra Malik," Region based Convolutional Networks for Accurate Object Detection and Segmentation", in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, 2016, pp. 142-158.
4. Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Danial Mollura, Ronald M Summers, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning", in IEEE Transactions on Medical Imaging ,vol. 35, Issue 5, May 2016, pp. 1285-1298.
5. Xuesong Zhang, Fei Yan, Yan Zhuang, Huosheng Hu, Chunguang Bu, "Using an Ensemble of Incrementally Fine-Tuned CNNs for Cross-Domain Object Category Recognition", in IEEE Access, vol. 7, March 2019, pp. 33822-33833.
6. Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, Xindong Wu, "Object Detection with Deep Learning: A Review", in arXiv preprint arXiv: 1807.05511, Apr 2019
7. P. N. Druzhkov, V. D. Kustikova, "A Survey of Deep Learning Methods and Software Tools for Image Classification and Object Detection," in Pattern Recognition and Image Analysis, vol. 26, 2016, pp. 9-15.
8. Chamarty Anusha, P. S Avadhani, "Object Detection Using Deep Learning", in International Journal of Computer Applications, vol. 182(32), Dec. 2018,pp. 18-22.
9. K. Simonyan and A. Zisserman "Very Deep Convolutional Networks for Large-Scale Image Recognition" in International Conference on Learning Representations 2015.
10. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke and Andrew Rabinovich "Going Deeper with Convolutions", in the Conference on Computer Vision and Pattern Recognition 2015.
11. Francois Chollet, "Deep Learning with Depthwise Seperable Convolutions" in the Conference on Computer Vision and Pattern Recognition, 2017.
12. Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks", in the Conference on Advances in Neural Information processing systems, 2012.

AUTHOR PROFILE

Mrs. Chamarty Anusha has received B. Tech degree from Jawaharlal Nehru Technological University, Andhra Pradesh and M.Tech from Andhra University, Visakhaptnam. The author is currently pursuing her Ph.D under Visvesvaraya Ph.D Scheme, Government of India from the Department of Computer Science and Systems Engineering, Andhra University, Visakhapatnam. Her main research work focuses on Object Detection using Deep Learning, Cyber Security and Machine Learning. She has published some research papers in various International Journals and attended numerous workshops and conferences during her career. The author has five years of experience in teaching field and dealt with various subjects in Computer Science like Computer Networks, Data Structures, Database designing etc.

Prof. P S Avadhani pursued M.Tech and Ph.D from IIT Kanpur. He is currently working as a Professor in the Department of Computer Science and Systems Engineering Andhra University, Visakhapatnam. His areas of interest are Cyber Security, Fuzzy Logic, Cyber Forensics, Computer Algorithms, Public Cryptography, and Data and Network Security. The author has written 6 books in Computer Science Engineering on various subjects. The author has published 52 papers in International Journals, 8 papers in National Journals, 38 papers in International Conferences and 9 papers in National Conferences. He has received several awards for his excellence in Teaching, Research and Administration like "State Best Teacher award", "Best Researcher award", "Distinguished Principal award" from the Government of Andhra Pradesh. The author had trained various PhD, M.Tech & B.Tech Engineering students during his 35 years of long teaching career.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)