International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-4, February 2020

Node Behavior Classification for Traffic Prediction in Optical Burst Switched Networks using Machine Learning

Deepali Bhawarthi , Girish Chowdhary

Abstract: Currently due to massive use of internet there is need of huge amount of bandwidth. The utilization of bandwidth can be managed up with optical burst switched networks. These networks cannot provide good QoS due to problems like wavelength contention and congestion problem. Also it is not necessary that contention in a network leads to congestion. It can be due to nodes behavior which affects the flow of traffic from source to destination. Hence there is a need to classify the traffic through the node at correct juncture to avoid congestion. This can be achieved using machine learning techniques. In this paper, support vector machine, AdaBoost classifier and Bagging classifier are evaluated .Experimental work is carried on Optical Burst Switched network dataset using 22 attributes which is available on UCI repository. The results show that bagging classifier performed better with accuracy of 95% in classifying the nodes behavior.

Keywords: Optical burst switched networks, Machine learning methods, Support vector machine, AdaBoost, Bagging

I. INTRODUCTION

In today's world there is need of huge amount of bandwidth due to intensive use of internet applications. In such cases it is very difficult to predict nature of internet traffic as it is bursty. So there is a need to ensure predefined QOS mechanism for such applications and hence there is need of new approaches. Wavelength Division Multiplexing (WDM) is a modern transport network technology, which utilizes transmission of bandwidth using optical fibers and has turn out to be observable select for backbone networks. It is used in many telecommunication industries [1, 3].Each fiber is allied with several communication channels and each of these channels are operative on different wavelength. Optical network provides with certain solutions through switching technology. It provides with 3 schemes namely optical circuit switching (OCS), optical packet switching (OPS) and optical burst switching (OBS).In OCS, a static communication is established during entire communication session avoiding bottlenecks due to optical electronic optical conversion (O-E-O) at intermediate nodes.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author

Deepali Bhawarthi*, Department of Computer Engineering, AISSMS College of Engineering, Pune, India. Email:deepali.ujlambkar@gmail.com

Dr Girish Chowdhary, School of Computational Sciences, SRTMU, Nanded, India. Email: girish.chowdhary@gmail.com

But it does not provide optimal solution for utilization of bandwidth problem due to bursty nature of traffic [4].In OPS, packets are switched optically without converting into electric signal. Packet loss issue was addressed using buffers. Optical packets are hold using fiber delay lines (FDLs).But to delay a single burst for 5s, it requires over a kilometer of fiber and hence was not flexible due to its expensive nature. [7], [8]. OBS has proved to be a good aspirant switching technique, which utilizes the advantages of both OPS and OCS [9], [10].OBS is designed as intermediate solution for OPS and OCS. In OBS, burst is the switching component which consists of collection of IP packets. There are two channels through which data and control packet are send respectively. First, a control packet is send to through the control channel, whose task is to reserve all the resources for the upcoming data burst. At every node the control packet is processed electronically. The data burst which is following control burst is transmitted transparently without conversion after some delay which is the offset time. The time gap between the transmission of data burst after the control burst is referred as offset time [11]. Without waiting for acknowledgement the OBS source node transmits data packet after offset time. The *offset time*, denoted as Ot, is the sum of total time taken by the control packet for processing at each intermediate node and is given by:

$$Ot = \sum_{h=1}^{n} \Delta h$$

Where, n is the total number of hops h and $\Delta h$is the delay at each hop such that $1 \leq h \leq n$. Statistical multiplexing is offered in OBS and OPS. But both these techniques lack buffering and use one way reservation hence are more prone to contentions. Contention losses do not indicate the condition of congestion always. Hence classification is needed to avoid false congestion identification which is achieved through machine learning (ML) techniques.

Following sections discussed the work carried out in this paper. In section 2, literature review is provided with different machine learning methods evaluated for prediction of channel. A comparative study on classifiers using dataset is performed to avoid congestion on a link. Section 3 described the proposed model, ensemble methods and performance metrics used in this study for evaluation. Results of experiments are presented in section 4 and section 5 provided the conclusion of this work.

II. RELATED WORK

From the past data the system behavior is learnt and future responses are estimated based on learned system model. This is unique capability of machine learning [14].

Such techniques are mostly used to solve issues like traffic predication, routing, congestion and contention control, resource management and QoS [4]. To enhance QoS, contentions are needed to resolve. Popular contention resolution techniques are namely optical buffering, deflection routing, wavelength conversion as well as segment discarding. There are various factors such as routing, wavelength opting criteria, nodes in the network, source to destination path length [16].Losses in OBS can be classified using Machine learning approaches. It is used to optimize the performance using example data or past experience. Machine learning techniques have two phases namely learning (training) phase and classification phase. By building a model from the former, estimates are captured in the training phase which is given as input to the classifier, which then classifies the data set [13].

ML approaches are basically classified into supervised, unsupervised and reinforcement learning categories. Depending on the feature set used as discrete or continuous supervised learning is classified as classification and regression task. It is based on the label input. A few algorithms to state in this category are K nearest neighbor, ANN, SVM [10].Unsupervised learning are bifurcated as clustering and algorithms like K means clustering, Principle component analysis (PCA) [16]. Unlike Supervised learning, reinforcement learning is based on agent's communication with the environment .It builds and updates ML model and needs no training. Q learning is the most popular technique used in this category [12].A gist of learning algorithms is shown in fig.1.

In [1], authors used both supervised and unsupervised machine learning technique is used to classify losses into clusters viz congestion and contention using heuristic methods. The accuracy of the classifiers was evaluated under different network conditions. In [2], authors proposed a detection approach which can classify different user's access patterns based on classification algorithm referred as sparse vector decomposition and rhythm matching. Authors in [5] proposed an approach for load balancing of traffic using alternative predefined paths. In [15], a global load-balancing contention resolution scheme is proposed and its performance is examined for both dynamic and static traffic. Another way to avoid contention is to implement a TCP-like congestion avoidance mechanism to regulate the burst transmission rate [2], [6], [7]. In this approach, the ingress edge nodes receive TCP ACK packets from egress edge nodes, calculate the most congested links, and reroute their traffic accordingly. The above proposed methodologies are related with traffic management schemes where the most congested link is managed by routing through alternative path which is predefined. But in real world scenario the traffic is bursty in nature and needs to be dynamically allocated to a wavelength to avoid such type of losses. The authors in [8] have presented a reinforcement method called Q learning deflection method approach to deal with congestion and scales along all nodes in the network. In [1], EM clustering is used as a representative of unsupervised learning and HMM for supervised learning to demonstrate the effectiveness and accuracy of the loss classification technique in different network scenarios and directs for loss classification which can be used to tune different applications under dynamic network scenarios. For more details on machine learning, the reader may refer to [13].

Classification deals with detection of harmful nodes which can be cause of congestion at the initial phase. This can lead to network performance deterioration. Many classification methods were analyzed and used by researchers .It addresses the area of research where accuracy is needed classify the nodes behavior. The purpose of this work presented in this paper is to apply a classification technique which helps for correct prediction of nodes behavior, which can lead to minimum delay and avoid congestion in OBS networks.

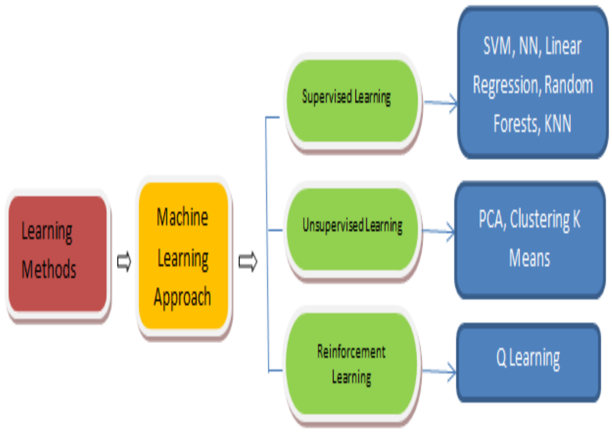

Fig.1 Learning Methods

III. MATERIAL AND METHODS

The proposed model describes the classification approaches using support vector machine (SVM), AdaBoost and bagging classifier. To validate the classifiers performance experimental setup was built on OBS network dataset. It is publically available on UCI repository. The proposed framework is designed using machine learning algorithms to classify the behavior of node broadly as blocking or nor blocking which can affect the network traffic. Three machine learning classifiers namely support vector machine, AdaBoost and bagging classifier are evaluated. The system is validated and evaluated through the performance metrics. The model works as follows:

1) Data preprocessing and feature selection.
2) Building a model to train our data using machine learning algorithms.
3) Comparison of results under different traffic scenarios of SVM, AdaBoost and bagging classifier

The following subsections briefly discuss the dataset and methods of the proposed work.

Dataset:

Dataset for OBS networks is available on UCI repository where various parameters are studied and are listed in table I. For classification this dataset was evaluated for presence of any detrimental node. The sample size of this dataset is 3225 with 22 attributes with some missing values.

There are 14 float features and 7 integer features with categorical feature. The target variables were assigned based on their processing behaviors like block, no block, NB-wait and NB-No block. The information about the features is described in table I.

Table- I: Attributes of OBS dataset

Attribute Name	Description
Node	Number of nodes in the network
Utilized Bandwidth Rate	Specifies the utilized bandwidth
Packet Drop Rate	specifies the drop rate of packet
Full_Bandwidth	Total bandwidth allocated to the network
Average_Delay_Time_Per_Sec	Average delay in seconds
Percentage_Of_Lost_Packet_Rate	packet loss in the network
Percentage_Of_Lost_Byte_Rate	Bytes lost in the network
Packet Received Rate	Number of packets received
Percentage of Used_Bandwidth	Total bandwidth used
Percentage of Lost_Bandwidth	Bandwidth wasted or unused in the network
Packet Size_Byte	Packet size
Packet_Transmitted	Number of packet transmitted
Packet_Received	Total number of packets received at the destination
Packet_lost	Total number of packets lost
Transmitted_Byte	Total bytes transmitted in the network
Received_Byte	Total bytes received in the network
10-Run-AVG-Drop-Rate	Average drop rate after 10 iterations
10-Run-AVG-Bandwith-Use	Average bandwidth used after 10 iterations
10-Run-Delay	Delay after 10 iterations
Node Status	Classification of nodes Viz. Block ,No Block ,Wait-NB,NB-No block
Throughput	Output
Class	Detection of type of node Viz NB-No Block, Block, No Block, NB-Wait

Data pre-processing and visualization:

In order to feed data into training and testing set to fit into machine learning model we need to preprocess data. It helps to analyze outliers and clean data. For this purpose we have used a box plot which is a graphical rendition of statistical data based on the minimum, first quartile, median, third quartile, and maximum. It contains a rectangle with graph extending from top to bottom [17]. The graph gives us information of numeric variables and also outliers if present. A shown in fig.2, few of the parameters like10-Run-Delay, Burst Loss, Average_Delay_Time_Per_Sec have outliers indicated as black dots. After checking the accuracy of these models the outliers are treated.

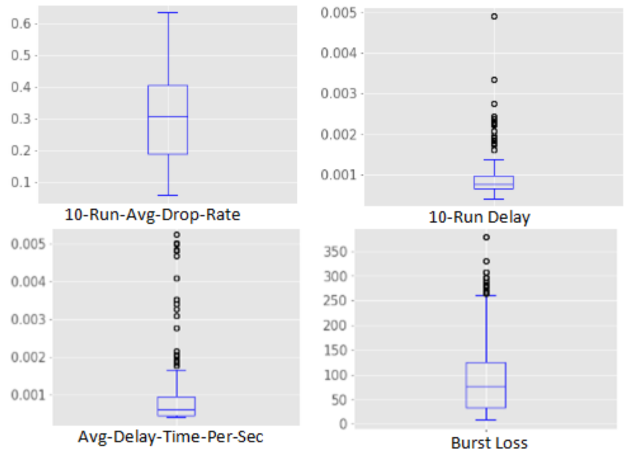

Fig.2 Outlier detection using whisker plot

Feature selection using PCA:

Now these features needs to be scaled .While retaining the variation of the dataset to maximum extend ,we need to reduce the dimensionality using principle component Analysis(PCA) in which new set of variables are formed using the old ones and are referred as principle components[18].

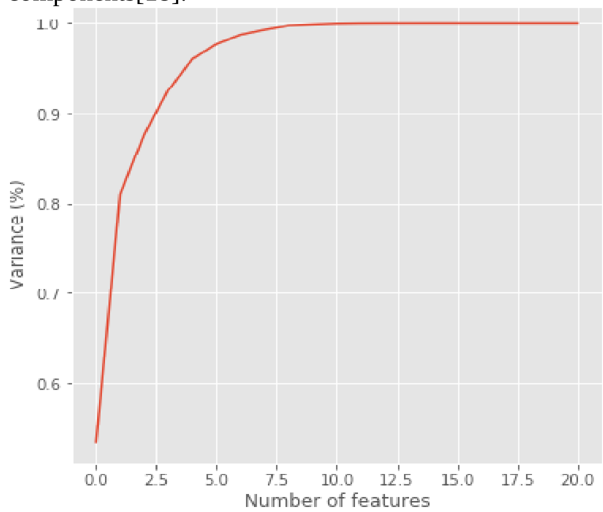

Fig.3 PCA Variance

From the above diagram, we conclude that we can reduce our dataset to 15 features. The values returned by PCA variance ratio are actually the eigenvalues. The cumulative sum of Eigen values is plotted which will help to identify the number of features which will preserve maximum variance.

Variance is plotted on Y-axis and numbers of features are plotted against X-axis [18].

Performance Evaluation Metrics:

The performance metrics like F1-score, Precision, Recall are used to classify nodes into classes as block, no-block, wait-Not behaving (NB) and NB-no block using parameters in table I for three classifiers support vector machine, AdaBoost classifier and bagging classifier. With this we can analyze the behavior of the node to predict the congestion in the network. For this the classifiers performance was checked with various evaluation metrics. The assessment of classification was done through following measures:

1) Accuracy: It is the ratio of number of correctly predicted occurrences to total occurrences in a given dataset.

2) Confusion matrix: Is used to describe the performance of classification model. In our model prediction is made between predicted classes and actual classes. The terms associated with confusion matrix are:

a) True positive (TP) – these are Correctly Identified means the predicted and actual output is true

b) False positive (FP) – Incorrectly Identified means the predicted output is true and actual output is false.

c) True Negative (TN) – Correctly rejected means actual and predicted output is false.

d) False Negative (FN) – Incorrectly rejected means actual output is true but predicted as false.

3) Recall: It is known as true positive rate. It is the amount of correct predictions i.e. positives compared to the total predictions throughout the data. It measures the proportion of actual correct predictions identified.

4) Precision: It is ratio of correctly identified positive predictions to the all positive predicted outcomes of the classifier

5) F1 score: Harmonic mean between recall and precision is F1 score and it ranges from 0 to 1.The models performance is rated by this measure .Greater the score, better is the performance of the model.

6) ROC-AUC curve: This measure helps to visualize the performance of multi classification problem. The probability measure is represented by receiver operating characteristics (ROC) and Area under Curve (AUC) represents degree of separability. It helps to discriminate amongst the classes. If the area under curve is higher, then the model is better. It is generally plotted against true positive values and false positive values [19]. AUROC for support vector machine and bagging classifier is shown in Fig.4 and Fig.5 respectively.

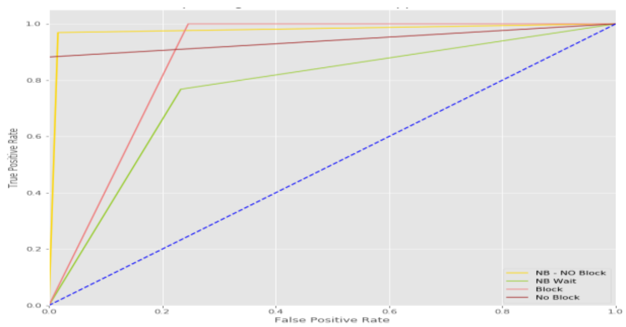

Fig.4. ROC for SVM

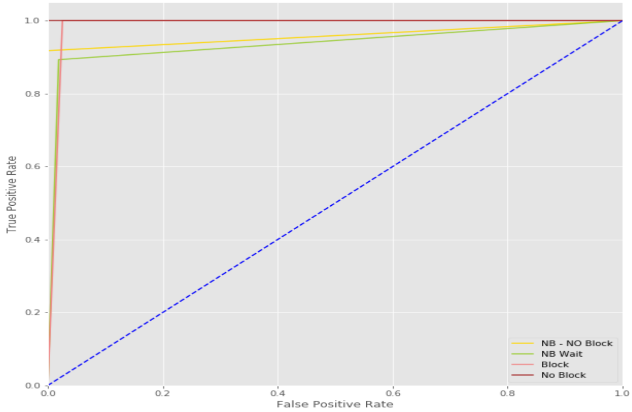

Fig.5 ROC for Bagging Classifier

IV. RESULT AND DISCUSSIONS

This section discuss about the performance evaluation of classifiers using Support vector machine, AdaBoost and Bagging classifier. First the dataset is divided into two sets as training and testing sets. Using proportional random sampling we ensure that the dataset contains sample of each type. The classifier is trained on first dataset and performance is checked on the second test dataset. This process is repeated and average value of test results is taken into consideration. Performance of classifiers is cross validated.

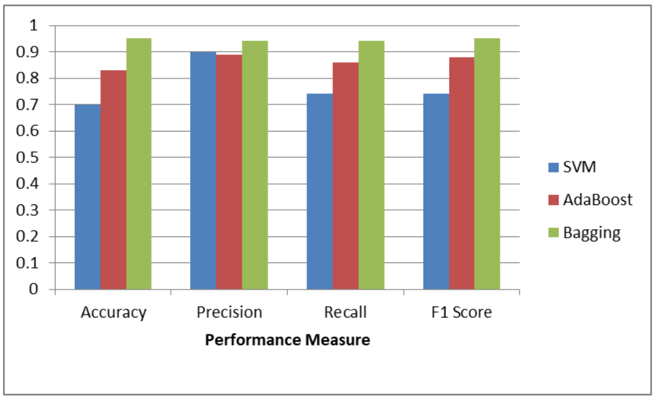

Fig.6 Performance evaluation of classifiers

Bagging classifier performs well with high accuracy, precision, recall and F1 score. The accuracy of bagging classifier was 95% with AdaBoost for 83% and SVM with 70%. SVM has poor performance as compared with other classifiers. Results show that bagging classifier has achieved better performance.

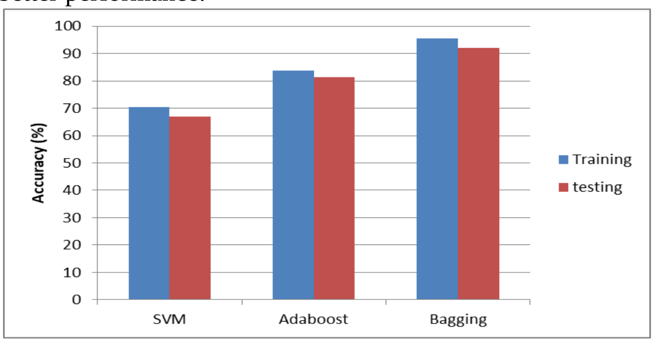

Fig.7. Performance evaluation of training and testing dataset

Figure shows the analysis of accuracy of training and testing data set which is used in this paper and infers that bagging classifier predicts accurate class labels as compared with SVM and AdaBoost classifier.

V. CONCLUSION

In this paper, we used SVM, AdaBoost and Bagging classifiers to classify the behavior of the node which can be probable cause of congestion in OBS networks, though it can utilize good amount of bandwidth. Due to use of high end applications on internet and its bursty nature traffic there is a need to acquaint with the behavior of node. For this purpose classification is proposed to classify the nodes depending on its behavior as no block, block, in Wait state or as a non-behaving-no block state using machine learning approach. Though OBS is a good contender to make huge utilization of bandwidth, but due to bursty traffic there are chances of congestion. So when the burst arrives at ingress node, it needs to reach the destination without delay. This is done through classification. The dataset was tested for three classifiers SVM, AdaBoost and bagging, out of which bagging classifier predicts better the behavior of the node based on the performance metrics like accuracy, AUC, Precision, Recall and F1 score. It predicts accuracy of 95% as compared with SVM and AdaBoost. The further enhancement for this work will be provided by using optimization methods to increase throughput by minimizing the burst loss probability due to congestion in OBS networks.

REFERENCES

1. A. Jayaraj, T. Venkatesh, and C. Siva Ram Murthy, “Loss Classification in Optical Burst Switching Networks using Machine Learning Techniques: Improving the Performance of TCP”, Senior Member, IEEE.IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, VOL. 26, NO. 6, AUGUST 2008
2. Q. Liao, H. Li, S. Kang, and C. Liu. ,”Application layer ddos attack detection using cluster with label based on sparse vector decomposition and rhythm matching”, Journal of Security and Communication Networks 2015;8(17): 3111–3120.
3. A. Detti and M. Listanti, , “Impact of Segments Aggregation on TCP Reno Flows in Optical Burst Switching Networks”, In: Proceedings IEEE INFOCOM, 2002
4. H. Wen, H. Song, L. Li, and S. Wang, “Load-balancing contention resolution in LOBS based on GMPLS”, In: Proceedings Fourth International Conference on Parallel and Distributed Computing, Applications and Technologies, Aug. 27-29 (2003), pp. 590-594.
5. J. Li, G. Mohan, and K.C. Chua, “Load Balancing Using Adaptive Alternate Routing in IP-over-WDM Optical Burst Switching Networks”, In: Proceedings SPIE Optical Networking and Communication Conference (OptiComm) 2003, Dallas, TX, vol. 5285, October (2003), pp. 336-345.
6. S.Y.Wang, “Using TCP congestion control to improve the performances of optical burst switched networks”, In: Proceedings IEEE ICC 2003, Anchorage, Alaska, USA, Vol. 2, May 11-15 (2003), pp. 1438-1442.
7. X. Cao, J. Li, Y. Chen, and C. Qiao, “Assembling TCP/IP Packets in Optical Burst Switched Networks”, Proceedings, IEEE GLOBECOM 2002, Taipei, Taiwan, November 17-21 (2002).
8. Haeri, Soroush & Thong, W.W.-K & Chen, Guanrong & Trajkovic, Ljiljana. (2013), “ A reinforcement learning-based algorithm for deflection routing in optical burst-switched networks“, Proceedings of the 2013 IEEE 14th International Conference on Information Reuse and Integration, IEEE IRI 2013. 474-481. 10.1109/IRI.2013.6642508.
9. Danish Rafique and Luis Velasco “Machine Learning for Network Automation: Overview, Architecture, and Applications [Invited Tutorial]”, Journal of Optical Communications and Networking, Vol. 10,Issue 10,pp. D126-D143 (2018)..
10. S. B. Kotsiantis, “Supervised machine learning: a review of classification techniques,” in Emerging Artificial Intelligence Applications in Computer Engineering—Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies, IOS, 2007, pp. 3–24.
11. S. G. Petridou, P. G. Sarigiannidis, G. I. Papadimitriou, and A. S. Pomportsis, “On the use of clustering algorithms for message scheduling in WDM star networks,” Journal of Lightwave Technol., vol. 26, no. 17, pp. 2999–3010, Sept. 2008.
12. C. J. C. H. Watkins and P. Dayan, “Q-learning,” Mach. Learn., vol. 8, no. 3, pp. 279–292, May 1992.
13. E. Alpaydin, “Introduction to Machine Learning”, 1st ed. MIT Press,2004
14. https://www.osapublishing.org/jocn/fulltext.cfm?uri=jocn-10-10-D126&id=396379
15. Boutaba, R., Salahuddin, M.A., Limam, N. et al. J Internet Serv Appl (2018) 9: 16. https://doi.org/10.1186/s13174-018-0087-2
16. Meiqian Wang, Shuo Li, Eric W. M. Wong and Moshe Zukerman, “Evaluating OBS by Effective Utilization”. IEEE Communications letters, Vol. 17, No. 3, March 2013
17. https://whatis.techtarget.com/definition/box-plot
18. https://www.dezyre.com/data-science-in-python-tutorial/principal-component-analysis-tutorial
19. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

AUTHORS PROFILE

Deepali Bhawarthi, is currently working with Department of Computer Engineering, AISSMS, College of Engineering, Pune, India. She has completed her master’s degree in Computer engineering from Pune University. She is pursuing her PhD in Computer Engineering from SRTMU, Nanded, India. Her research interest areas are Optical networks, Data Science, Machine Learning, Soft Computing and Big Data Analytics.

Dr Girish Chowdhary , is currently working as Professor and Dean, School of Computational Sciences, SRTMU, Nanded, India. His area of interest includes Optical networks, Machine learning, Security, etc. He has presented and published several research papers in reputed journals and conferences. He is also a reviewer for many reputed journals.