Integrating Machine Learning Approaches in SDN for Effective Traffic Prediction Using Correlation Analysis

Bhuvaneswari Balachander[1](✉), Manivel Kandasamy[2], Venkata Harshavardhan Reddy Dornadula[3], Mahesh Nirmal[4], and Joel Alanya-Beltran[5]

[1] Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Kanchipuram, India
bhuvaneswari@saveetha.com

[2] Unitedworld School of Computational Intelligence, Karnavati University, Gandhinagar, Gujarat, India
manivel.k@karnavatiuniversity.edu.in

[3] IT & ITES, Startups Mentoring Society, Tirupati, India

[4] PREC, Loni, India

[5] Electronic Department, Universidad Tecnológica del Perú, Lima, Peru
C18121@utp.edu.pe

Abstract. The study shows that numerous academic researchers are utilizing machine learning and artificial intelligence approaches to regulate, administer, and run networks, as a result of the recent explosion in interest in these fields. In contrast to the scattered and hardware-centric traditional network, Software Defined Networks (SDN) are a linked and adaptive network that offers a full solution for controlling the network quickly and productively. The SDN-provided network-wide information may be used to improve the efficiency of traffic routing in a network environment. Using machine learning techniques to identify the fewest overloaded path for routing traffic in an SDN-enabled network, we investigate and demonstrate their application in this study. These years have seen an increase in the number of researchers working on traffic congestion prediction, particularly in the field of machine learning and artificial intelligence (AI). This study topic has grown significantly in recent years on account of the introduction of large amounts of information from stationary sensors or probing traffic information, as well as the creation of new artificial intelligence models. It is possible to anticipate traffic congestion, and particularly short-term traffic congestion, by analyzing a number of various traffic parameter values. When it comes to anticipating traffic congestion, the majority of the studies rely on historical information. Only a few publications, on the other hand, predicted real-time congestion in traffic. This study presents a comprehensive summary of the current research that has been undertaken using a variety of artificial intelligence approaches, most importantly distinct machine learning methods.

Keywords: Software Defined Network (SDN) · Machine Learning (ML) · Artificial intelligence · Traffic · Quality of Service (QoS) · Information · Datasets

V. E. Balas et al. (Eds.): ICETCE 2022, CCIS 1591, pp. 611–622, 2022.
https://doi.org/10.1007/978-3-031-07012-9_51

1 Introduction

Conventional networks are naturally dispersed, and they route or forward signals using hop-based routing strategies. The path with the fewest hops between any pair of nodes is picked, and all traffic is routed down that path. This method of transferring data worked well when networks were first created, but as they gotten bigger and utilization, issues like as latency began to emerge, strangling the lines and requires the interaction inefficient. Due to the substantial control overhead imposed by communicating modified network information in a distributed fashion, routing techniques such as hop-based approaches do not include the current amount of network problems into their route computations. However, this often results in wasteful network resource consumption, since signals may be routed along a little long corridor that is less crowded than the congested least steps path, resulting in improved traffic stress sharing in the networks [1].

SDN (Software Defined Networking) is a revolutionary networking architecture that allows for centralized management of network assets by combining control and information planes [2]. It acts as a customization link between network elements and the centralized administrator. The centralized controller offers a worldwide view of the whole network, allowing for more versatility in regulating and running the network to effectively fulfill the required Quality of Service (QoS) need. Conventional networks lack such a coherent worldwide perspective of the network. Machine Learning is now widely employed in a wide range of applications & industries. Using labeled samples, supervised machine learning techniques are used to anticipate future occurrences [3]. The method infers a suitable function to produce effective findings by studying a known experimental dataset. Unsupervised machine learning methods, on the other hand, are utilized when no previous knowledge is able to instruct the algorithms.

It attempts to derive assumptions from the datasets in attempt to characterize underlying commonalities and traits in the unorganized information. Semi-supervised machine learning methods train using a mix of labeled and unlabeled input. This ensemble may be used to significantly increase acquisition accuracy. Machine learning techniques that use reinforcement learning create actions and alter them in response to punishments or incentives. This technology enables machines & software programs to autonomously find the optimal behavior in a given scenario in order to optimize their efficiency. Due to their dispersed and dumb nature, machine learning methods are difficult to integrate and implement in conventional networks to govern and run networks. SDN opens up new avenues for incorporating intelligence into networks [4].

2 Objective

The research aimed to fulfill the following objectives:

1. To study the Machine Learning
2. To study the Software Defined Networking (SDN)
3. To study the Traffic Prediction
4. To study the Traffic Classification
5. To study Optimized Routing on the Basis of Traffic Classification

3 Methodology

Machine learning & artificial intelligence to regulate networks. SDNs transform traditional networks into linked, complex networks that provide a full network management platform. SDN's network-wide data may help enhance traffic routing. For traffic monitoring in an SDN network, we examine and show the application of machine learning approaches. In current history, traffic congestion prediction has become a major topic in Artificial Intelligence (AI) and machine learning (ML). New AI models and visualization of information from permanent sensors and investigation devices have expanded this research area. Different traffic parameters are assessed to predict short-term traffic congestion.

4 Machine Learning

Machine Learning is defined as the branch of research that investigates how computers may learn without even being specifically programmed in a certain way. This might be viewed of as a more formalized version of an earlier casual judgment. Machine learning is defined as: When a computer programme learns from experiences E with regard to a group of tasks T as well as a performance measure P, it is said to have learned from experience E if its performance at tasks in T, as measured by P, increases over time. It is possible to describe machine learning (ML) as the research journal of techniques & statistical modelling that computerized networks use to do a certain job without utilizing detailed instructions, rather depending on trends & assumptions [5] as a subfield of machine learning.

Table 1 shows the results of the study. Machine Learning can be divided into 4 main categories: supervised, unsupervised, semi-supervised, & reinforcement learning. Supervised learning is the most common kind of machine learning. Detailed descriptions of various algorithms are provided in the next section [6] (Fig. 1).

Table 1. The results of the study

Learning method	Description
Supervised machine learning	To build a supervised learning model, you provide the system with learning algorithm, that is, inputs and anticipated outcomes/outputs. With fresh data, it can forecast using the learned connection. The training data is labelled. Supervised learning issues fall into two categories: regression & categorization. While in regression issues, the theory predicts continuously outcomes, classification models generate discontinuous outcomes. Many supervised learning methods exist [7]
Unsupervised machine learning	Machine learning without supervision is the absolute opposite of supervised learning. The system is provided with a set of inputs information without their matching outputs in this paradigm. The algorithm groups the information regarding the relationships between the information's "features" [8]. k-means & self-organizing mappings are 2 main instances of such techniques
Semi-supervised machine learning	Semi-supervised learning is a kind of learning that falls between supervised and unsupervised. In this paradigm, a portion of the input information has associated outputs (labelled), whereas the rest does not (unlabeled). 'Transductive' SVM & graph-based approaches are examples of such techniques [9]
Reinforcement learning	It is the challenge of determining how individuals should conduct in a given environment in order to maximize progressive rewards [10]. An agent interacts with its surroundings in order to learn the appropriate behaviors to do in order to maximize long-term benefit

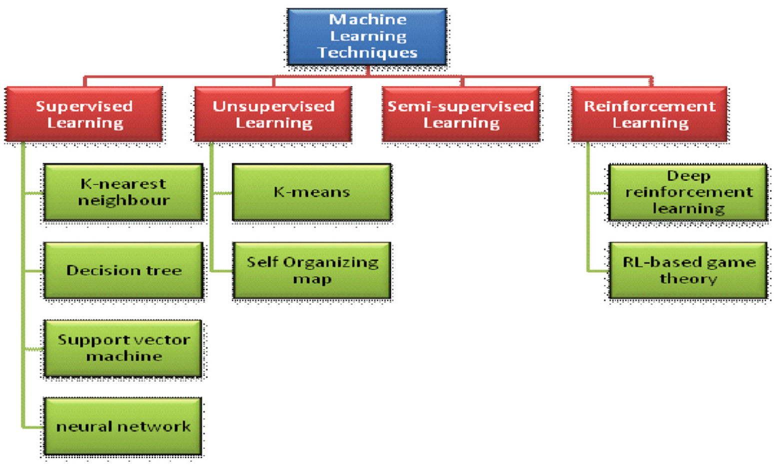

Fig. 1. Techniques for machine learning

5 Software Defined Networking (SDN)

The internet as we understand it is mostly built of the connectivity of old or heritage networking, which is what we call the backbone of the internet. Traditional networks are comprised of specialized network equipment such as networking equipment, which is liable for both processes of moving choices and transmitting data throughout the network. These network devices route traffic based on just a limited understanding of the network's configuration. The network structure of the Legacy system is scattered. Network devices are composed of many thousands of code lines that are exclusive and do not allow for any degree of customization. The intricacy of conventional networks [11] is one of their most significant drawbacks. The demands placed on networks have changed throughout the years. In response to these evolving needs, the more complicated the architecture of network devices becomes, as well as the increased cost of this equipment. It is also difficult to apply new regulations in old networks because of their established infrastructure. Adding an extra device or introducing a new service in an already big network would need a thorough setup that would include numerous network devices being used in tandem. This will take a significant amount of time & financial assets (Fig. 2).

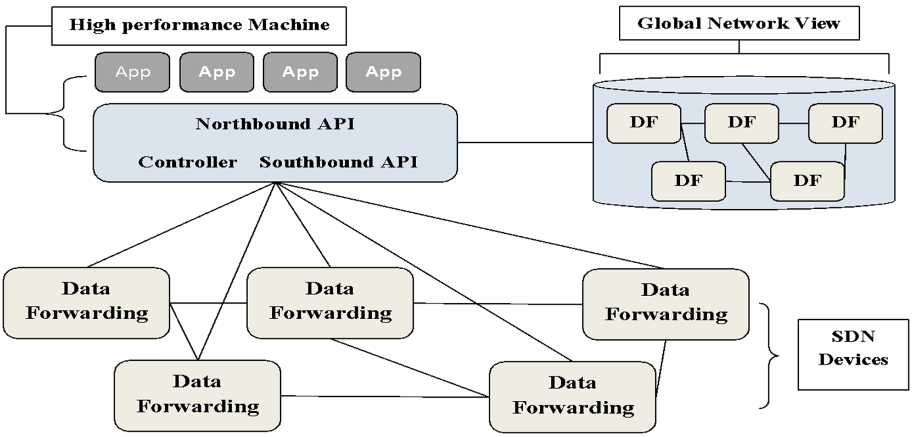

Fig. 2. Design of Software Defined Networking (SDN)

Because the usage of Internet of Things devices is growing at such a fast pace, conventional networks are plainly falling short of the mark to satisfying today's network needs. In current history, there has been a significant increase in interest in software-defined networking. Many studies have been conducted in recent years to determine how we might leverage the flexibility provided by SDN to control traffic and enhance computer connectivity, particularly in light of the increasing usage of Internet of Things (IoT) devices. SDN is essentially the splitting of the network switch into stages [12] and is not complicated. This centralized component, which manages networks gadgets such as switches, shields the network's intellect away from the rest of the network. When compared to the restrictions of conventional networks, SDN provides more network flexibility.

In classical networks, a router or switch is considered to be a singular object that comprises both the control plane (the brain) that makes decisions and the information plane or transmitting plane that is accountable for the forwarding of packets of information. With the introduction of SDN, there seems to be a decoupling between the control plane & the information planes. In other words, there is a specialized central controller that is independent of the rest of the system and that manages the transmitting switches. The transmitting mechanisms on the information plane are completely incompetent. Their intellect might be limited or non-existent if they are built in this way. There are several benefits to this concept [1]. It provides network managers with the ability to programme networks in order to meet more specialized or customized demands. In addition, as compared to traditional networking, where the switches only get a limited view of the network, the central unit has a worldwide perspective of the network, which is advantageous (Fig. 3).

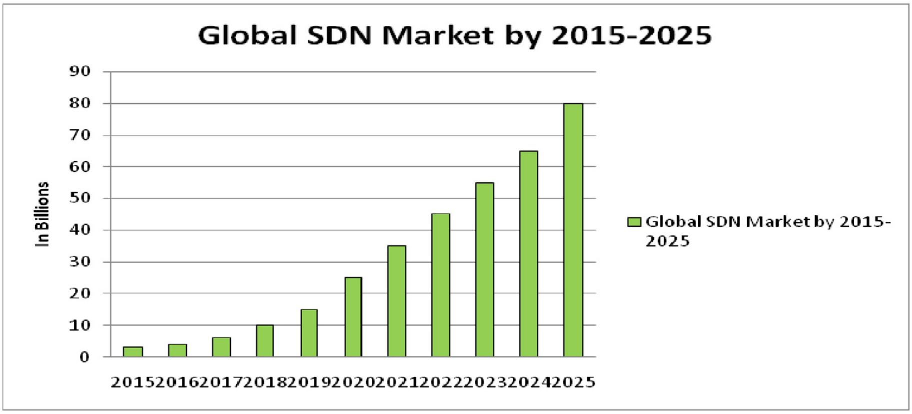

Fig. 3. Global Software Defined Networking (SDN) market by 2015–2025

6 Traffic Prediction

Network traffic prediction is critical in the management & administration of today's more sophisticated and diversified networks, and it is becoming more important. It includes predicting future traffic volumes, and has typically been handled via the use of time series forecasting (TSF). The goal of TSF is to develop a regression framework that is suitable of drawing an adequate link amongst projected volume of traffic and recently known volume of traffic [13]. Current TSF models for traffic prediction may be divided into two categories: statistically analytical methods and supervised machine learning (ML) models. Advanced auto - regressive integrating movement averages (ARIMA) methods are often used in statistical methods, but supervised neural networks (supervised NNs) are used to train the vast number of traffic prediction algorithms in practice.

Generalized auto-regressive (AR) or movement averages (MA) methods are used in combination to conduct auto-regression on differences or "stationarized" information in the ARIMA model, which is a common strategy for TSF. As a result of the fast expansion of networks and the rising sophistication of network traffic, classic TSF models appear to be undermined, resulting in the development of more powerful machine learning models [14]. Other than traffic volume, attempts have been made in latest days to minimize inefficiency and/or increase correctness in traffic predictions by including information from flows, which are distinct from traffic volume, into the forecast process. Numerous traffic prediction algorithms that make use of machine learning are described in the following subsections, and their results are summarized in Table 2.

Table 2. Numerous traffic prediction algorithms

ML technique	Application	Dataset	Features	Output
Supervised:- SVR	Prediction of link load in ISP networks (TSF)	Internet traffic gathered at an ISP network's POP (N/A)	Link stress measured on a timeframe	Link load prediction
Supervised:- MLP-NN with different training algorithms (GD, CG, SS, LM, Rp)	Network traffic prediction (TSF)	1000 points datasets (N/A)	Previous observations	Traffic volume that is expected
Supervised:- NNE trained with Rp	Predictions of link stress and traffic flow in ISP networks (TSF)	Traffic on a transatlantic connection aggregating traffic in the ISP backbone (N/A) SNMP traffic information from two ISP networks	Estimated traffic volume in the last few minutes multiple days	Traffic volume that is expected
Supervised:- MLP-NN	Prediction of route bandwidth availability from beginning to finish (TSF)	NSF Turgid datasets (N/A)	Load maximum, minimum, and average values measured in the last 10 s to 30 s	The amount of capacity that will be available on an end-to-end link in a future era
Supervised:- KBR-LSTM-RNN	Using historical internet flow data to predict future traffic volume (regression)	Over a 24-week timeframe, network traffic volumes and flowing count data were recorded every 5 min (public)	Flow count	Traffic volume that is expected

7 Traffic Classification

Traffic categorization is crucial for network administrators in order to carry out a broad variety of network administration and managerial functions efficiently and effectively. Capacity management, security & detection techniques, quality of service & service segmentation, efficiency measurement, & resource management are just a few of the capabilities available. For instance, a corporate network operator may also want to prioritize traffic for business-critical apps, recognize an unknowable traffic for outlier detection, or undertake workload characteristics in required to design efficient material managerial strategies that meet the efficiency and resource specifications of a variety of apps [15].

It is necessary to be able to properly correlate network traffic with pre-defined classes of concern in order to perform traffic classification. It is possible that these categories of interests will be classified as classes of applications (e.g., e-mail), classes of apps (e.g., Skype, YouTube, or Netflix), or classes of services [16]. An application or class of apps that have the same QoS criteria are grouped together as a class of service, for example, based on Quality of Service (QoS). As a result, it is feasible that apps that seem to function differently from one another are really part of the same level of services [17].

In generally, network traffic categorization algorithms may be divided into four main groups based on specific port, packet content, host behavior, and flow characteristics [18]. The traditional technique to traffic categorization simply links applications with port numbers that have been published with the Internet Assigned Numbers Authority (IANA). But, because it is no longer the de facto standard, and because it does not offer itself to learning owing to the need for minimal lookup, it is not included in the scope of this study. Furthermore, it has been evidenced that simply relying on dock numbers is ineffectual, owing in large part to the use of vibrant port bargaining, tunneling, and the mishandling of port assigning numbers to the well apps for the purpose of distorting traffic and ignoring firewalls, among other factors. Although several classifiers use port numbers in combination with other strategies [19] to increase the effectiveness of the traffic classifiers, this is not always the case. Following that, we will explore the different traffic categorization algorithms that make use of machine learning, which will be summarized in Table 3.

Table 3. Different traffic categorization algorithms

ML technique	Datasets	Characteristics	Classes	Evaluation	
				Settings	Result
Supervised RF, Boost, Boost	Exclusive: business network	Packet dimension (from 1 to N packets), packet timestamp (from 1 to N packets), inter-arrival duration (from 1 to N packets), source/destination MAC, source/destination IP, source/destination port, flow length, packet count and byte count are all variables to consider	YouTube, Skype, Face book, Web Browsing	N = 5	RF: Accuracy 73.6–96.0% SGBoost: Accuracy 71.2–93.6% XGBoost: Accuracy 73.6–95.2%
Semi-supervised Laplacian-SVM	Univ. network is a unique	unpredictability of packet length, packet transmission duration, due to impaired, destination port, packages to react, min package length from destinations to sources, packet inaction degrees from transmitting data	Voice/video conferencing, streaming, large-scale data transmission, and more services are available	N = 20, Laplacian-SVM parameters $\lambda = 0.00001-0.0001$, $\sigma = 0.21-0.23$	Accuracy > 90%
Supervised k-NN, Linear-SVM, Radial-SVM, DT, RF, Extended Tree, AdaBoost, Gradient-AdaBoost, NB, MLP	KDD	Connection numbers, connection percentages, login status, error rate	Attack types	Dynamic classification and features collection selection	Accuracy = 95.6%

8 Optimized Routing on the Basis of Traffic Classification

In this part, we will look at works that have utilized machine learning to categorize traffic and then gone on to identify the ideal route based on traffic class that has been identified. [20] makes use of machine learning and software-defined networking to enhance scheduling & control in information centers. There were two stages to the framework. The initial phase entailed applying Machine Learning to categories the traffic at the network's edges in order to identify patterns. Using the categorization results in combination with its worldwide perspective of the network, the centralized SDN controller is able to deploy efficient routing options. This study did not go into detail on which methods and processes would be utilized for categorization & routing, nor did it specify which methods and procedures will be used for classification and routing. Another piece of work, which combines traffic categorization with route optimization, is provided in this publication. Within a software-defined network (SDN) context, they presented a Deep Neural Network (DNN) classifiers for network traffic categorization [21].

A virtual network function (VNF) is incorporated into the design in order to lessen the strain on the SDN. When a flow comes, the SDN switches routes the flow in accordance with the flow rules that have been configured. If a package does not have a flow rule associated with it, it makes a query to the controllers. Depending on the network architecture, the controllers determine the best path to take for the flow to be routed and then installs these patterns in the SDN switches in addition, the controller creates a flow rule that directs a duplicate of the packet to a port where the VNF will be listening for it. Traffic is captured and characteristics are extracted for use in traffic categorization using the trained deep neural network (DNN). The SDN controller receives the specified class data, which is marked in the DSCP field and transmitted to it. A route that meets the QoS criteria of the defined traffic class is now sought by the SDN controller, which then assigns major value to the resulting flows in all switches along the route, giving them preference over the flows that were originally installed [22]. This paper proposes a strategy for improving connection speeds, however it does not suggest any optimization techniques that may be employed by the SDN controller after classifications in order to find the best routing for the network traffic. In order to route the recognized packet, the controllers simply utilize its standard routing method, which is based on the worldwide perspective.

9 Conclusion

In this study, we have suggested a machine learning approach for congestion conscious traffic prediction in software defined networks. The suggested model learns updated network information at various periods of time and integrates that knowledge for picking least overloaded channel for traffic management. The study and predictions of traffic has been a focus of continuing research in several sub-fields of computer networks. Countless group of experts have been developed an efficient network traffic technique for the analysis and the predictions of network traffic and also, we evaluated the optimum routing on the basis of traffic categorization.

References

1. Kreutz, D., Ramos, F.M., Verissimo, P.E., Rothenberg, C.E., Azodolmolky, S., Uhlig, S.: Software-defined networking: a comprehensive survey. Proc. IEEE **103**(1), 14–76 (2015). https://doi.org/10.1109/jproc.2014.2371999
2. Crucianu, M., Boujemaa, N.: Active semi-supervised fuzzy clustering. Pattern Recogn. **41**(5), 1834–1844 (2008). https://doi.org/10.1016/j.patcog.2007.10.004

3. Panwar, V., Sharma, D.K., Kumar, K.V.P., Jain, A., Thakar, C.: Experimental investigations and optimization of surface roughness in turning of en 36 alloy steel using response surface methodology and genetic algorithm. Mater. Today: Proc. (2021). https://doi.org/10.1016/j.matpr.2021.03.642
4. Abar, T., Ben Letaifa, A., El Asmi, S.: QoE enhancement over DASH in SDN networks. Wirel. Pers. Commun. **114**(4), 2975–3001 (2020). https://doi.org/10.1007/s11277-020-07513-w
5. Rimal, Y.: Machine learning prediction of Wikipedia. Int. J. Mach. Learn. Netw. Collab. Eng. **3**(2), 83–92 (2019). https://doi.org/10.30991/ijmlnce.2019v03i02.002
6. Yu, F., Huang, T., Xie, R., Liu, J.: A survey of machine learning techniques applied to software defined networking (SDN): research issues and challenges. IEEE Commun. Surv. Tutor. **21**(1), 393–430 (2019). https://doi.org/10.1109/comst.2018.2866942
7. Mahrishi, M., Morwal, S., Muzaffar, A.W., Bhatia, S., Dadheech, P., Rahmani, M.K.I.: Video index point detection and extraction framework using custom YoloV4 Darknet object detection model. IEEE Access **9**, 143378–143391 (2021). https://doi.org/10.1109/ACCESS.2021.3118048
8. Holzinger, A.: Introduction to machine learning and knowledge extraction. Mach. Learn. Knowl. Extr. **1**(1), 1–20 (2017). https://doi.org/10.3390/make1010001
9. Jain, A., Pandey, A.K.: Multiple quality optimizations in electrical discharge drilling of mild steel sheet. Mater. Today Proc. **4**(8), 7252–7261 (2017). https://doi.org/10.1016/j.matpr.2017.07.054
10. Henderson, P., Bellemare, M., Pineau, J.: An introduction to deep reinforcement learning. Found. Trends Mach. Learn. **11**(3–4), 219–354 (2018). https://doi.org/10.1561/2200000071
11. Meena, G., et al.: Traffic prediction for intelligent transportation system using machine learning. In: 2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE), pp. 145–148 (2020). https://doi.org/10.1109/ICETCE48199.2020.9091758
12. Kapre, P., Shreshthi, R., Kalgane, M., Shekatkar, K., Hande, Y.: Software defined networking based intrusion detection system. Int. J. Recent Trends Eng. Res. **3**(5), 475–480 (2017). https://doi.org/10.23883/ijrter.2017.3252.m8yed
13. Jain, A., Pandey, A.K.: Modeling And optimizing of different quality characteristics in electrical discharge drilling of titanium alloy (grade-5) sheet. Mater. Today: Proc. **18**, 182–191 (2019). https://doi.org/10.1016/j.matpr.2019.06.292
14. Troshin, A.V.: Machine learning for LTE network traffic prediction. Infokommunikacionnye tehnologii 400–407 (2019). https://doi.org/10.18469/ikt.2019.17.4.06
15. Zúquete, A.: Traffic classification for managing applications' networking profiles. Secur. Commun. Netw. **9**(14), 2557–2575 (2016). https://doi.org/10.1002/sec.1516
16. Yu, C., Lan, J., Xie, J., Hu, Y.: QoS-aware traffic classification architecture using machine learning and deep packet inspection in SDNs. Procedia Comput. Sci. **131**, 1209–1216 (2018). https://doi.org/10.1016/j.procs.2018.04.331
17. Padmanabhan, V., Seshan, S., Katz, R.: A comparison of mechanisms for improving TCP performance over wireless links. ACM SIGCOMM Comput. Commun. Rev. **26**(4), 256–269 (1996). https://doi.org/10.1145/248157.248179
18. Akodkenou, I., Soule, A., Salamatian, K.: Traffic classification on the fly. ACM SIGCOMM Comput. Commun. Rev. **36**(2), 23–26 (2006). https://doi.org/10.1145/1129582.1129589
19. Jain, A., Yadav, A.K., Shrivastava, Y.: Modelling and optimization of different quality characteristics in electric discharge drilling of titanium alloy sheet. Mater. Today Proc. **21**, 1680–1684 (2020). https://doi.org/10.1016/j.matpr.2019.12.010
20. Cost efficient resource management framework for hybrid job scheduling under geo distributed data centers. Int. J. Mod. Trends Eng. Res. 4(9), 77–84 (2017). https://doi.org/10.21884/ijmter.2017.4282.yq9ne

21. Zhang, C., Wang, X., Li, F., He, Q., Huang, M.: Deep learning-based network application classification for SDN. Trans. Emerg. Telecommun. Technol. **29**(5), e3302 (2018). https://doi.org/10.1002/ett.3302
22. Deebalakshmi, R., Jyothi, V.L.: Smart routing based on network traffic classification techniques and considerations. Int. J. Public Sect. Perform. Manag. **5**(1), 1 (2019). https://doi.org/10.1504/ijpspm.2019.10016149