Optimal Feature Selection to improve Vehicular Network Lifetime

Sakshi Garg[1 [0000-0001-6467-9065]], Deepti Mehrotra[2 [0000-0001-5752-9800]] and Sujata Pandey[3 [0000-0003-2669-8323]] and Hari Mohan Pandey[4 [0000-0002-9128-068X]]

[1,2,3]Amity University, Noida, Uttar Pradesh, India
[4]Bournemouth University, United Kingdom
[1]sakshijyotigarg@gmail.com
[2]mehdeepti@gmail.com
[3]spandey@amity.edu
[4]profharimohanpandey@gmail.com

Abstract. The advancement of the Internet of Things (IoT) leads to the ascent of the need to develop a protocol for Low power and Lossy Networks (LLNs). The IETF ROLL working group then proposed an IPv6 routing protocol called RPL in 2012. RPL is in demand because of its adaptability to geography changes and its ability to recognize and dodge loops. Although RPL in the recent past was only used for IoT networks. But, contemporary studies show that its applicability can be extended to vehicular networks also. Thus, the domain of the Internet of Vehicles (IoV) for RPL is of significant interest among researchers. Since the network becomes dynamic when RPL is deployed for vehicular networks, the heterogeneous network suffers from extreme packet loss, high latency and repeated transmissions. This lessens the lifetime of the organization. The thought behind this article is to simulate such a dynamic environment using RPL and identify the principal features affecting the network lifetime. The network setup is simulated using the Cooja simulator, a dataset is created with multiple network parameters and consequently, the features are selected using the Machine Learning (ML) technique. It is inferred from the experiment that increasing PDR and reducing EC will improve the overall network lifetime of the network.

Keywords: RPL, IoV, Network Lifetime, PDR, EC, ML.

1 Introduction

Internet of Things (IoT) [1] network permits devices and people to collectively exchange information. IoT is further considered a dynamic network which leads to the evolution of a new domain altogether called the Internet of Vehicles (IoV) [2]. IoV par-adigm includes Vehicle-to-Roadside (V2R), Vehicle-to-Vehicle (V2V), Vehicle-to-In-frastructure (V2I) and Vehicle-to-Communication network (V2C). LLNs based on the IoV network face multiple constraints like energy, memory, reliability, heterogeneity, etc. Therefore, it is necessary to optimize LLN issues concerning IoV. Since several studies [3][4][5] are available in the literature that has targeted LLN-based IoT network and only a few works [6][7] can be located that have explored IoV for LLNs, it still remains contemporary distress.

RPL was introduced in 2012 by the IETF ROLL working group to tackle these LLN issues [8]. RPL to date is used for IoT networks [9] but recent state-of-art [10] has shown its implementation for IoV networks too. RPL within IoV requires development to ad-dress heterogeneity, power Quality of Service (QoS) and versatility. Multiple scientists have proposed RPL advancement to further develop network execution [11][12] con-cerning IoV networks but to enhance the network performance it is vital to analyze the network at its core.

In this article, an optimal feature selection is proposed to identify the chief features affecting the network lifetime. It is evident from the literature [13] that the higher the network lifetime, the higher the network performance. It is also well established [14] that the network lifetime can be boosted by increasing the Packet Delivery Ratio (PDR) and reducing Energy Consumption (EC). Arbitrarily proposing enhancements and optimization to RPL within the IoV network are of little significance on contrary to propos-ing solutions with optimally chosen network parameters. This motivated us to simulate a mobile environment using the RPL protocol, collect various parameters affecting the network and finally chose the optimal features to improve the network lifetime. The selection is performed on the collected dataset by using Principal Component Analysis (PCA) for dimensionality reduction and Extra Tree Classifier (ETC) to identify those reduced features. The accuracy of the proposed selection is also tested using Machine Learning (ML) technique to justify the efficiency of the selected parameters. Network lifetime is improved by dropping EC and rising PDR. Thus, the parameters touching these two factors are calculated in this study. The results justify that the major contrib-uting factors affecting PDR are Throughput (THPT), Control Traffic Overhead (CTO) and Total Latency (TL). The prime parameters responsible for EC are Expected Trans-mission Count (ETX), Sent Packets (SP) and TL. It can also be seen that TL is a common factor for both PDR and EC. The accuracy of this selection is 91.756% and 90.116% respectively.

The paper is additionally organized as: the literature study is summed up in Section II. Section III presents the specifics of the gathered dataset. Section IV discusses the technique. Section V infers the results and in section VI the conclusion with possible future directions is conferred.

2 Literature Study

The literature suggests the necessity to improve network lifetime to improve network performance using RPL [15]. Such enhancements will prove beneficial for both IoT and IoV networks. The authors of the paper [16] proposed the use of ETX and EC to enhance network lifetime for IoT networks by utilizing the Network Interface Average Power metric (NIAP). The parameters chosen to improve network lifetime are not based on scientific calculations but merely on observations. Also, it is limited to the IoT network implementation, while in this paper, the parameters are optimally chosen based on the ML strategies. In paper [17], the authors proposed a new objective function based on energy, the number of siblings and ETX. The results showed an improvement in network lifetime by lowering delay at 95% of PDR. Their proposal is for IoT networks and metric consideration to form objective function is freehold.

Authors [18] advised the use of Mobility Energy and Queue Aware-RPL (MEQA-RPL) to increase network lifetime. Their results show improvement in network lifetime but the results could be better if the metric parameter were optimally selected, which is suggested in our study. Authors in this review paper [19] clearly highlighted that increase in PDR and lower EC can elevate network lifetime significantly. Paper [20] also utilized the trade-off between PDR and latency, load balancing and transmission performance to maintain a higher network lifetime. Authors of the paper [21] recommended the use of the ML technique to optimally select features for predictive modelling. Our paper is so conceptualized to select the most optimal network parameters for the vehicular networks and later the accuracy of the selected features is predicted by modelling it using the LR technique to justify the preciseness of the feature selection process.

These studies intended at improvising the RPL performance and mostly optimizing RPL by improving objective function metrics but very few studies have discussed network execution of RPL. In this article this fall is handled by proposing optimal feature selection to enhance network lifetime for vehicular networks.

3 Dataset

We have accumulated the data by simulating the environment over the Cooja simulator on Contiki OS. This dataset contains multiple factors that may contribute to the network statistics. The experimental setup includes various parameters as shown in Ta-ble 1. The data is collected by increasing the number of nodes at continuous intervals to consider scalability as one of the network parameters. Each experiment has been re-capped at least 5 times and then an average of those values is taken to reduce any biases. So, the original dataset measures 30 X 16 (6 node intervals X 5 times repetition X 16 parameters). Further, Fig. 1 shows the a short illustration of preprocessed dataset.

A	B	C	D	E	F	G	H	I
Nodes	**HC**	**RTM**	**ETX**	**EC**	**PDR**	**CTO**	**THPT**	**SP**
10	1.580	1037.417	32.136	1.667	52.432	28904	0.570	49260
20	1.478	1051.614	32.119	2.260	51.546	95096	1.616	139650
30	1.697	1069.849	33.666	2.209	50.725	122970	2.187	188987
40	1.882	1206.850	38.141	2.425	47.619	182845	3.802	328502
50	1.520	1074.142	40.223	2.276	38.043	200934	4.008	346262
60	1.688	1345.975	41.944	2.895	29.458	289742	4.734	409023

Fig. 1. An illustration of the collected dataset

The chief parameters of the dataset are explained briefly ahead:

- **Nodes**: They represent the density of the simulated network.
- **Routing Metric (RTM):** A unit calculated by the RPL algorithm to select or reject a path for data transmission.
- **Expected Transmission Count (ETX):** Total number of transmissions performed o send a packet from source to destination.
- **Energy Consumption (EC)**: Total power consumed by a mote during the life-time of the network.
- **Packet Delivery Ratio (PDR)**: A ratio of the total number of packets delivered to the total number of packets sent from source to destination.
- **Control Traffic Overhead (CTO)**: The sum of control messages created by the nodes in the network.
- **Throughput**: The rate of successful packet delivery over the network channel.
- **Sent Packets**: Total number of packets sent from source to destination.
- **Received Packets**: Total packets received during the network lifetime.
- **Lost Packets**: Total packets lost during network lifetime.
- **Total Latency**: Total round trip time the data packet takes to travel.
- **Simulation Time**: Total time the experiment was performed.

4 Methodology

To predict the important features of the network, the steps are shown in Fig. 2 as follows. The model is developed using a python script on a jupyter notebook.

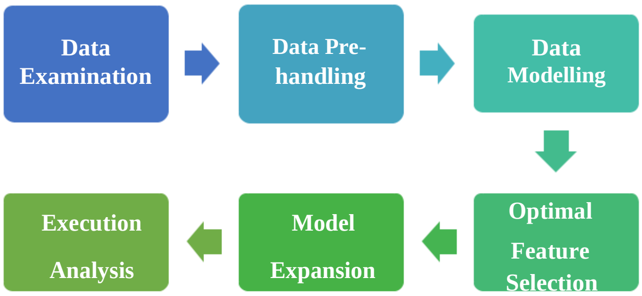

Fig. 2. Steps for optimal feature selection

4.1 Issue Declaration

It is certain that the extension in network execution and organization lifetime with growing scalability, will upgrade the execution and QoS assuming it is expected and anticipated beforehand. Limited exploration of vehicular networks within IoV prompts the need to evaluate the chief parameters affecting the network. It is apparent from the literature that improving network lifetime can significantly improve network QoS and performance. Hence, it is crucial to identify prime parameters that impact network per-formance. So, it breeds the necessity to select optimal features that can enhance network performance and provide QoS with great precision.

4.2 Proposed Model

Data Exploration. The data is explored by simulating the dynamic environment using the RPL protocol. To incorporate mobility in RPL, the BonnMotion tool is exploited to generate mobility and the Random Waypoint mobility model is applied to ensure a ver-itable representation of vehicular network traffic. Further, the data is collected by per-forming diverse experiments and some parameters are obtained using the Wireshark network analysis tool like Total Latency, Throughput and Control Traffic Overhead. The parameters set for the data collection in the Cooja simulator are shown in Table 1 and Table 2.

Table 1. Simulation Detail using Cooja

Index	Parameters
Random Seed	123,456
Start-up Mote Delay	65 s
Radio Messages	6LoWPAN with pcap analyzer
Model of Propagation	UDGM with constant Distance Loss
Mote	Sky
Nodes’ Variation	10, 20, 30, 40, 50, 60
TX Ratio	100%
RX Ratio	50%
INT Range	90 m
TX Range	45 m
Total Simulation Time	24 h

Table 2. Details for BonnMotion tool

Index	Parameters
Nodes' Variation	10, 20, 30, 40, 50, 60
Mobility Model	Random WayPoint
Area X * Y	100 m^2
Clustering Range	10 m
Pause time	20 s
Number of WayPoint	6
Environment	Mobile

Data Pre-processing and Cleaning. The primary data is cleaned of any Null values. Multiple readings of the same experiments are averaged to obtain holistic values of the parameters. Any parameter that is not contributing to the prediction of the target pa-rameter (Network Lifetime) is dropped. Like, the simulation time for all the experi-ments is the same, hence its standard deviation is zero, and therefore it is dropped from the dataset. Redundant parameters like CV RP, CV SP, and CV LP are also ignored. An example of the refined dataset is displayed in Fig. 3.

```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        from sklearn.ensemble import ExtraTreesClassifier

In [2]: df = pd.read_csv('PCA.csv')

In [6]: ndf=df.drop(['ST','CV RP','CV SP','CV LP'], axis=1)

In [7]: y = ndf['PDR']
        X = ndf.drop('PDR', axis = 1)

In [8]: X.head()

Out[8]:
```

	Nodes	HC	RTM	ETX	EC	CTO	THPT	SP	TL	RP	LP
0	10	1.580	1037.417	32.136	1.667	28904	0.570	49260	5.267	25828.216	23431.784
1	20	1.478	1051.614	32.119	2.260	95096	1.616	139650	4.992	71984.536	67665.464
2	30	1.697	1069.849	33.666	2.209	122970	2.187	188987	12.181	95862.971	93124.029
3	40	1.882	1206.850	38.141	2.425	182845	3.802	328502	18.338	156429.524	172072.476
4	50	1.520	1074.142	40.223	2.276	200934	4.008	346262	19.978	131730.109	214531.891

Fig. 3. Refined data illustration

Data Modelling. The data is then modelled using Extra Trees Classifier to identify the significant features with respect to PDR and EC.

Optimal Feature Selection. Our objective is to escalate network lifetime, which is evident from the literature that it can be accomplished by increasing PDR and reducing EC. For this, we need to identify critical features affecting PDR and EC. Evaluation of varied features in the network increases the network complexity. Calculating all those parameters to estimate the target variable Network Lifetime is needless. Thus, dimensionality reduction is done using Principal Component Analysis (PCA) to select the most optimal features concerning PDR and EC.

Model Development. The obtained features are then modelled using Linear Regression to predict the accuracy of the model for PDR and EC separately.

Performance Analysis. The accuracy of the model is tested for PDR and EC both.

This article commends the need to increase PDR and reduce EC to increase Network Lifetime as a solution to decide the execution and upgrade it for better QoS and unri-valled execution. The utilization of ML strategy is the fundamental idea behind culti-vating this model, because of ML's ability to deal with gigantic data. Meanwhile the vehicular organization is dynamic, the size of information can't be forecasted. Subse-quently, this model purposes the multivariate Linear Regression strategy by picking ideal elements to create the model with great precision.

5 Results and Discussions

The primary dataset is modelled and important features impacting PDR and EC are plotted using the Extra Trees Classifier. The graph obtained is reflected in Fig. 4. and Fig. 5.

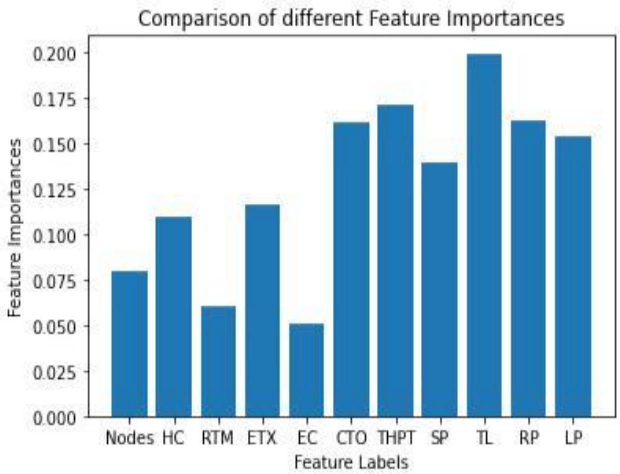

Fig. 4. Feature importance with respect to PDR

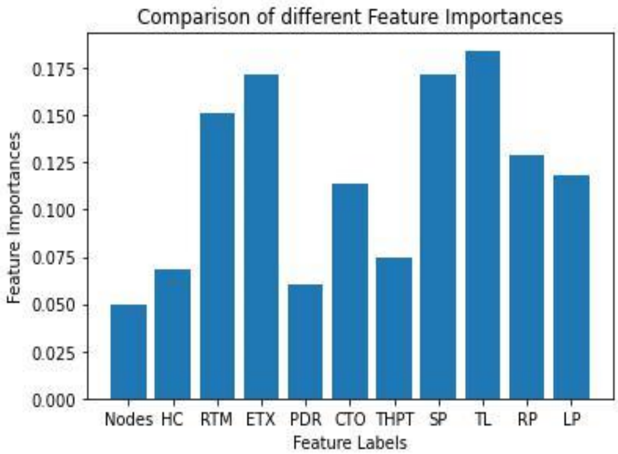

Fig. 5. Feature importance with respect to EC

Fig. 4. shows that TL is the most imperative feature and EC is the least affecting the PDR of the network. However, Fig. 5. shows that TL is the most vital feature and the scalability of nodes is least disturbing the EC of the network. Further, the dataset has numerous parameters and some features might not be essential for PDR or EC. Evaluating compound features repeatedly will increase the complexity of the network. Hence, it is obligatory to reduce features and find the most contributing ones. The dimension-ality reduction of the dataset is performed using PCA for PDR and EC both. Most con-tributing factors are identified out of multiple factors using PCA taking 95% of the components of the dataset parameters. The results reveal that there are three most im-portant features that impact PDR and EC significantly. It can be seen in Fig 6. and Fig. 7.

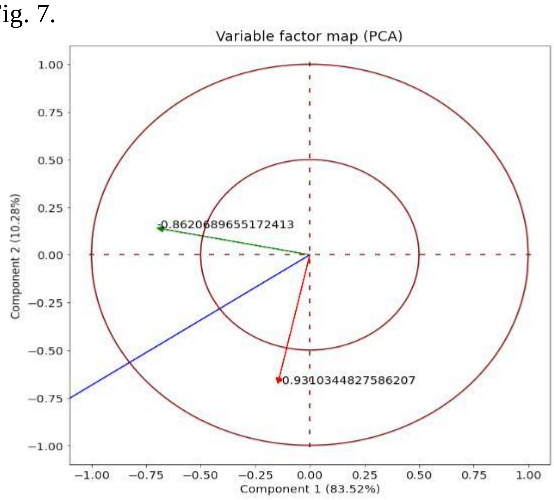

Fig. 6. Extracted features for PDR using PCA variable map

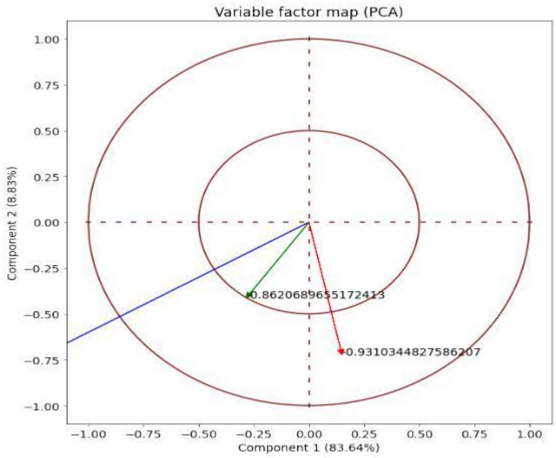

Fig. 7. Extracted features for EC using PCA variable map

The most important extracted features for PDR and EC are listed in Table 3.

Table 3. Optimally selected feature

Network Lifetime	Factors affecting PDR and EC
Increase PDR	1. Total Latency (TL) 2. Throughput (THPT) 3. Control Traffic Overhead (CTO)
Reduce EC	1. Total Latency (TL) 2. Sent Packets (SP) 3. Expected Transmission Count (ETX)

The accuracy of the selection is predicted using a machine learning technique called Linear Regression (LR) and it is seen that optimally selected features for PDR give 91.756% accuracy while the accuracy score for EC is 90.116%.

This table fairly clarifies the use of the recommended optimally selected features for evaluating the network performance and the key parameters influencing the network lifetime for vehicular networks.

6 Inference and Future Directions

This paper distinctly underscores the necessity to find optimal features for the IoV network. Predictive analysis is performed using ETC, PCA and LR strategies of ML to select the most optimal parameters for the vehicular network to upsurge network life-time. The dataset is first simulated by recreating the dynamic environment using the RPL protocol, mobile nodes and RWP mobility model. Then, the dataset is modelled using ETC to find the importance of features for PDR and EC. Further, the dataset is condensed using PCA to reduce dimensionality and thereby network complexity. The dataset is collected using the collect view in the Cooja simulator and processing the simulation log files over the Wireshark network analyzer. The obtained feature through PCA fit into the LR model to obtain the accuracy of the prediction. The precision score for the chosen features is anticipated to legitimize the effectiveness of the proposed parameters to upgrade network lifetime for vehicular networks. The precision for cho-sen features for PDR and EC is impartially perfect to legitimize its utilization for future vehicular organizations. The typical accuracy score for PDR is 91.756%, while the typ-ical exactness score for EC is 90.116%. This idea has direct applications for smart city organizations, Electronic Toll Collection organizations, and so forth. In future, this work can be stretched out for genuine information and other mobility models like Ran-dom Walk (RWK), Nomadic, Manhattan Grid, etc. to analyze the behavior of optimally selected features on network lifetime concerning those models.

References

1. Safaei, B., Mohammadsalehi, A., Khoosani, K. T., Zarbaf, S., Monazzah, A. M. H., Samie, F., ... & Ejlali, A. (2020). Impacts of mobility models on RPL-based mobile IoT infrastructures: An evaluative comparison and survey. *IEEE access*, *8*, 167779-167829.
2. Garg, S., Mehrotra, D., Pandey, H. M., & Pandey, S. (2021). Accessible review of inter-net of vehicle models for intelligent transportation and research gaps for potential future directions. *Peer-to-Peer Networking and Applications*, *14*(2), 978-1005.
3. Bendouda, D., & Haffaf, H. (2019, June). QFM-MRPL: Towards a QoS and Fault Man-agement based of Mobile-RPL in IoT for mobile applications. In *2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 354-359). IEEE.
4. Vaezian, A., & Darmani, Y. (2022). MSE-RPL: Mobility Support Enhancement in RPL for IoT Mobile Applications. *IEEE Access*.
5. Zarzoor, A. R. (2022). Enhancing IoT performance via using Mobility Aware for dy-namic RPL routing protocol technique (MA-RPL). *International Journal of Electronics and Telecommunications*, *68*(2), 187-191.
6. Garg, S., Mehrotra, D., Pandey, S., & Pandey, H. M. (2021). Network efficient topology for low power and lossy networks in smart corridor design using RPL. *International Journal of Pervasive Computing and Communications*.
7. Kumar, A., & Hariharan, N. (2020, February). Enhanced mobility based content centric routing in rpl for low power lossy networks in internet of vehicles. In *2020 3rd Interna-tional Conference on Intelligent Autonomous Systems (ICoIAS)* (pp. 88-92). IEEE.
8. Winter, T., Thubert, P., Brandt, A., Hui, J. W., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, J.P. & Alexander, R. K. (2012). RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks. *rfc*, *6550*, 1-157.

9. Darabkh, K. A., Al-Akhras, M., Ala'F, K., Jafar, I. F., & Jubair, F. (2022). An Innovative RPL Objective Function for Broad Range of IoT Domains Utilizing Fuzzy Logic and Multiple Metrics. *Expert Systems with Applications*, 117593.
10. Garg, S., Mehrotra, D., Pandey, H. M., & Pandey, S. (2022). Static to dynamic transition of RPL protocol from IoT to IoV in static and mobile environments. *Cluster Computing*, 1-16.
11. Seyfollahi, A., Moodi, M., & Ghaffari, A. (2022). MFO-RPL: A secure RPL-based rout-ing protocol utilizing moth-flame optimizer for the IoT applications. *Computer Stand-ards & Interfaces*, *82*, 103622.
12. Aboubakar, M., Kellil, M., Bouabdallah, A., & Roux, P. (2020, April). Using machine learning to estimate the optimal transmission range for RPL networks. In *NOMS 2020-2020 IEEE/IFIP Network Operations and Management Symposium* (pp. 1-5). IEEE.
13. Sennan, S., Ramasubbareddy, S., Nayyar, A., Nam, Y., & Abouhawwash, M. (2021). LOA-RPL: novel energy-efficient routing protocol for the internet of things using lion optimization algorithm to maximize network lifetime.
14. Cyriac, R., & Karuppiah, M. (2020). A comparative study of energy retaining objective functions in rpl for improving network lifetime in iot environment. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Sci-ence)*, *13*(2), 159-167.
15. Garg, S., Mehrotra, D., & Pandey, S. (2022). A Study on RPL Protocol with Respect to DODAG Formation Using Objective Function. In *Soft Computing: Theories and Appli-cations* (pp. 633-644). Springer, Singapore.
16. Pereira, H., Moritz, G. L., Souza, R. D., Munaretto, A., & Fonseca, M. (2020). Increased network lifetime and load balancing based on network interface average power metric for RPL. *IEEE Access*, *8*, 48686-48696.
17. Moradi, S., & Javidan, R. (2020). A new objective function for RPL routing protocol in IoT to increase network lifetime. *International Journal of Wireless and Mobile Compu-ting*, *19*(1), 73-79.
18. Durai, S. (2021). MEQA-RPL: A Mobility Energy and Queue Aware Routing Protocol for Mobile Nodes in Internet of Things.
19. Ghanbari, Z., Navimipour, N. J., Hosseinzadeh, M., Shakeri, H., & Darwesh, A. (2022). The applications of the routing protocol for low-power and lossy networks (RPL) on the internet of mobile things. International Journal of Communication Systems, e5253.
20. Jamalzadeh, M., Maadani, M., & Mahdavi, M. (2022). EC-MOPSO: an edge computing-assisted hybrid cluster and MOPSO-based routing protocol for the Internet of Vehicles. Annals of Telecommunications, 1-13.
21. Garg, S., Mehrotra, D., Pandey, S., & Pandey, H. M. (2021, November). A Convergence Time Predictive Model using Machine Learning for LLN. In 2021 IEEE 8th Uttar Pra-desh Section International Conference on Electrical, Electronics and Computer Engi-neering (UPCON) (pp. 1-5).