International Journal of Advances in Applied Sciences (IJAAS)
Vol. 14, No. 4, December 2025, pp. 1433~1443
ISSN: 2252-8814, DOI: 10.11591/ijaas.v14.i4.pp1433-1443

Forecasting internet traffic patterns for the campus Metro-E network using a hybrid machine learning model

Norakmar Arbain[1], Murizah Kassim[1,2], Darmawaty Mohd Ali[1], Shuria Saaidin[1]
[1]Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
[2]Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia

Article Info

Article history:

Received Feb 24, 2025
Revised Oct 30, 2025
Accepted Nov 4, 2025

Keywords:

Campus Metro-E network
Hybrid machine learning
Internet traffic forecasting
Network efficiency
Traffic management strategies

ABSTRACT

Complex traffic patterns lead to crucial campus Metro-E network management and resource allocation. This paper presents an internet traffic forecasting by pre-processing data to offer better bandwidth quality of service (QoS). Eight (8) campuses' traffic data were analysed for modelling predictions using statistical analysis. A Metro-E campus network presents four (4) locations: A, E, F, and H have is a strong correlation between inbound and outbound traffic, with correlation values between 0.4547 and 0.5204. As the inbound traffic increases, outbound traffic tends to rise as well. Conversely, locations B, C, and G have weak correlations, indicating more independent traffic patterns. Data outliers were found for locations C and F, where unusual traffic spikes require further network exploration and show key trends in traffic data. Descriptive statistics reveal notable differences, with H has the highest average traffic at about 75 Mbps, while C has the lowest at around 30 Mbps. Location F shows the greatest traffic fluctuation with a standard deviation of 0.4076, whereas Location G has very little fluctuation with a standard deviation of 0.0240. Overall, this pre-process data is use to combine machine learning (ML) to improve prediction abilities for better bandwidth management and real-time handling in digital campus environments.

Corresponding Author:

Murizah Kassim
Institute for Big Data Analytics and Artificial Intelligence, Universiti Teknologi MARA
40450 Shah Alam, Selangor, Malaysia
Email: murizah@uitm.edu.my

1. INTRODUCTION

In the context of today's digital landscape research, predicting internet traffic has become a crucial aspect. This involves estimating data volume on a network over a specific timeframe, which is essential for maintaining smooth operations. Accurate forecasting aids in managing networks, controlling congestion, and efficiently allocating resources, especially in campus Metro-E networks [1]. The rapid expansion of digital platforms highlights the necessity for accurate internet traffic forecasting, particularly as educational institutions increasingly rely on online resources and services [2]. As more users access educational content simultaneously, networks frequently experience unpredictable traffic spikes, resulting in congestion, latency, and a diminished user experience [3], [4]. Consequently, effective traffic forecasting is essential for the proactive management of network resources [5]. Predicting traffic flow patterns can enhance user satisfaction by managing bandwidth during surges. This review assessed existing internet traffic prediction models, highlighting their limitations and identifying areas for future research. A crucial aspect of developing these models is scalability and computational efficiency, as accuracy relies on the resources required for analyzing large datasets used by many institutions and users [6], [7]. Scalability refers to the model’s capacity to handle increasing amounts of data and its ability to adapt to any network growth without a corresponding degradation in performance [8], [9]. A key challenge faced by many traditional models is that they often exhibit a trade-off between accuracy and computational resources. While more complex algorithms, particularly those based on deep learning approaches, have shown improved accuracy in forecasting, they often require substantial computational power, which may not be readily available in all institutional settings. Challenges in performance management arise due to increased user density, high-bandwidth applications, and fluctuating demand, often resulting in network congestion that can degrade quality of service (QoS) through latency, packet loss, and reduced throughput [8], [10]. Many institutions use reactive network management, which addresses performance issues only after they arise. This limits the ability to anticipate demand and allocate resources effectively. While descriptive analytics can analyze historical data, it cannot predict future traffic changes. Predictive analytics with machine learning (ML) provides a better solution. Models like seasonal autoregressive integrated moving average (SARIMA) identify long-term patterns, while long short-term memory (LSTM) networks manage nonlinear dependencies. Combining these methods in a hybrid approach can enhance forecasting accuracy [11]–[13].

Predictive analytics in campus Metro-E networks offers benefits due to unique traffic patterns shaped by academic calendars, exams, events, and irregular user behaviors. However, research has largely focused on single-model frameworks, highlighting a gap in the development of hybrid predictive frameworks for campus or multi-campus networks [14], [15]. Predictive modeling of internet traffic in academic networks often relies on single methods like SARIMA, which handles seasonality and trends but struggles with sudden nonlinear changes. LSTM networks capture complex patterns but may overlook recurring cycles typical in academic settings, such as semester schedules. Few studies have developed hybrid frameworks for multi-campus higher education, leading to less accurate traffic predictions. A hybrid approach could improve forecasting accuracy and robustness by addressing both seasonal regularities and nonlinear bursts [16], [17]. A popular area of research involves developing hybrid ML models that combine the strengths of various algorithms [18]–[20]. Hybrid models combine time-series forecasting with ML algorithms, such as support vector machines and neural networks, to achieve high accuracy and efficiency in predicting internet traffic. Combining autoregressive integrated moving average (ARIMA) with neural networks can enhance predictions during peak usage periods. In education, user behavior has cyclical patterns, highlighting the need for efficient algorithms to optimize resources. Techniques like feature selection and dimensionality reduction can improve computation speed with minimal performance loss, which is essential for a campus Metro-E network requiring fast response times [21]–[23].

Internet traffic prediction with a hybrid ML model encompasses a variety of algorithms, ranging from statistical linear models to nonlinear ML methods. Traditionally, many researchers have relied on statistical techniques, such as time series analysis and regression models, to forecast traffic flow, striving for optimal accuracy in their results [24]. However, the advent of ML has significantly transformed traffic prediction capabilities, as these advanced algorithms can effectively model and learn from complex, nonlinear relationships within the data [25], [26]. Researchers in [27] have demonstrated the effectiveness of various ML approaches in achieving high accuracy levels in predicting traffic flows, thereby enhancing the overall efficiency of traffic management systems and contributing to better urban planning. Traditional statistical models, particularly time-series models, are commonly used for traffic forecasting by analyzing historical data for patterns. To address their limitations, researchers are developing hybrid models that combine statistical techniques with advanced ML algorithms [28], [29]. The ARIMA-LSTM model combines ARIMA's time series forecasting strengths with LSTM's ability to learn from sequences and manage long-term dependencies, enhancing the capture of both linear and non-linear patterns in traffic data. Similarly, the ARIMA-convolutional neural network (CNN) hybrid integrates ARIMA's linear trend and seasonal analysis with CNN's feature extraction, resulting in improved accuracy and insights compared to traditional models [30]. Time-series models are statistical techniques that analyze data points collected or recorded at specific time intervals. Numerous studies have applied ARIMA models to predict internet traffic. For instance, research has shown that ARIMA can effectively forecast traffic in both local area networks (LANs) and wide area networks (WANs). In studies conducted by Saha *et al.* [31], [32], ARIMA was used to predict traffic patterns based on real traffic datasets from various high-speed traffic data. Similarly, another study by Wang *et al.* [33] used an enhanced SARIMA model to analyze traffic in cellular zones within a residential community. SARIMA adds seasonal components to the traditional ARIMA to handle periodic fluctuations in traffic, making it particularly useful for predicting internet traffic in environments with clear seasonal trends [34]. SARIMA enhances forecasts by incorporating seasonal factors, offering greater accuracy during variations. However, it complicates parameter estimation with added seasonal parameters and still faces limitations, such as reliance on linear relationships, making it challenging to adapt to sudden changes in traffic patterns [35], [36].

Researchers in [37] indicate that LSTMs often outperform traditional models, such as ARIMA and exponential smoothing, in terms of predictive accuracy. Hybrid models offer accuracy and reliability but require extensive preparation and training time. Most research focuses on internet traffic prediction in wired LAN and mobile networks, with a gap in studies on campus Metro-E network traffic as shown in Table 1.

Table 1. Research gaps in hybrid ML models for internet traffic prediction

References (authors, year)	Hybrid model	Traffic type/context	Advantages	Limitations/research gaps
Saha and Haque [18] (2023)	Wavelet+ensemble ML	Internet traffic under distribution shifts	Decomposes signal; hybrid model improves out-of-distribution generalization over standalone models	Still performance drop under shift; limited generalization across different types of distribution changes.
Shi *et al.* [38] (2021)	NN-ARIMA (MLP/NN followed by ARIMA)	Network-wide traffic (flow, speed, occupancy)	Captures nonlinear patterns via NN; models residuals with ARIMA to refine accuracy	Residuals may still contain structure; choice of NN and ARIMA tuning remains ad hoc.
Saha *et al.* [39] (2024)	ConvLSTMTransNet (CNN+LSTM+Transformer)	High-speed port internet telemetry (time-series)	Captures spatial and temporal dependencies; ~10 % better accuracy vs. RNN/LSTM/GRU baseline	Not tested on multivariate scenarios or under online/adversarial settings.
Shao *et al.* [40] (2022)	CEEMDAN+PSO-LSTM (decomposition+PSO optimized LSTM)	Network traffic time series	Decomposes and denoises signals; PSO optimizes model training, yielding improved prediction	Complex setup; PSO may overfit; lacks validation across varying network environments.
Su *et al.* [41] (2024)	Lightweight hybrid attention+CNN	5G network traffic prediction	Efficient feature learning via attention; computationally lightweight via depth wise separable convs.	Generalization to other network types or scales not validated; complexity still exists, albeit reduced.
Norakmar (2025)	SARIMA+LSTM	Campus Metro-E network	Descriptive analysis and statistical models with hybrid predictive algorithms for campus Metro-E networks.	-

2. METHOD

2.1. Internet traffic pattern analysis for campus Metro-E network

This study preprocesses Metro-E internet traffic data to improve forecasting accuracy and resource allocation. By integrating hybrid ML models, it enhances network responsiveness and effectiveness. As research progresses, developing efficient models will be essential for managing campus Metro-E network traffic and ensuring high-quality user service. The analytical process in Figure 1 starts with collecting traffic flow data, followed by preprocessing for consistency, and descriptive statistical analysis to reveal insights like correlation and seasonality. Predictive modeling with SARIMA and LSTM techniques forecasts traffic demand, while distribution analysis assesses underlying patterns. The forecasting results help optimize bandwidth and enhance QoS on the campus network.

2.2. Internet traffic at campus Metro-E network

Ethernet technology began with LANs but has evolved into a top choice for metropolitan area networks (MANs). It offers good prices, easy management, various services, and low costs. In campuses, metro ethernet networks (MENs) connect different enterprise LANs across urban areas and work well with WANs from telecom providers. MENs use a strong fiber-optic infrastructure and Ethernet for communication, allowing for fast data transfer and reliable performance for high-bandwidth applications. Understanding traffic patterns helps institutions use resources wisely, manage congestion, and address maintenance needs, ensuring efficient network operations for the future. As depicted in Figure 2, the campus Metro-E network provisions bandwidth at an impressive rate of 10 Gbps for wired connections, while wireless connections through Unifi achieved 2 Gbps. Additionally, the evaluated traffic flow throughput reaches 700 Mbps, comprising 500 Mbps of inbound traffic and 200 Mbps of outbound traffic, thereby providing a clear representation of the network's operational capabilities and performance metrics.

2.3. Data collection and scope

Internet traffic flow has been monitored across various campus Metro-E network environments, with inbound and outbound throughput data collected in Mbps. The dataset underwent thorough preprocessing to enhance quality, including cleaning missing values, handling outliers, and normalizing data for consistency. Table 2 summarizes variable parameters for each location, offering a comprehensive analysis of internet traffic characteristics. This study analyzes data from eight campuses (A to H). It involved preprocessing to clean missing values and normalize data. Key metrics assessed network traffic, including correlation of flows, mean traffic volume, and measures of central tendency and dispersion, revealing variability across locations.

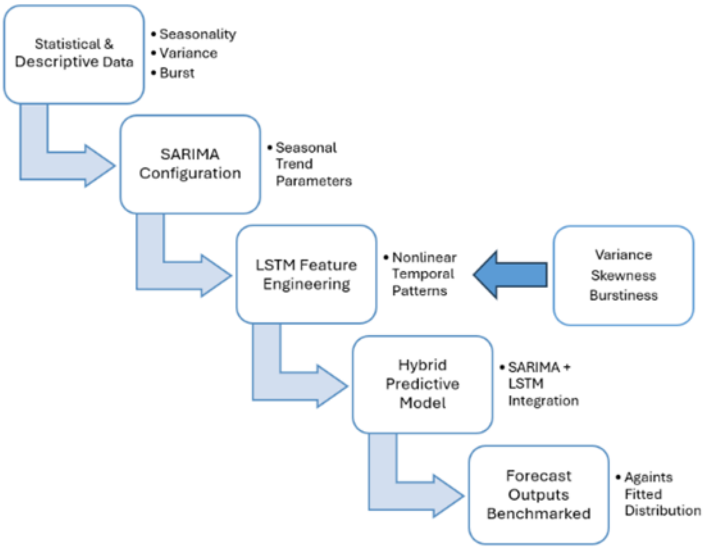

Figure 1. Analytical process for internet traffic forecasting in campus Metro-E Network

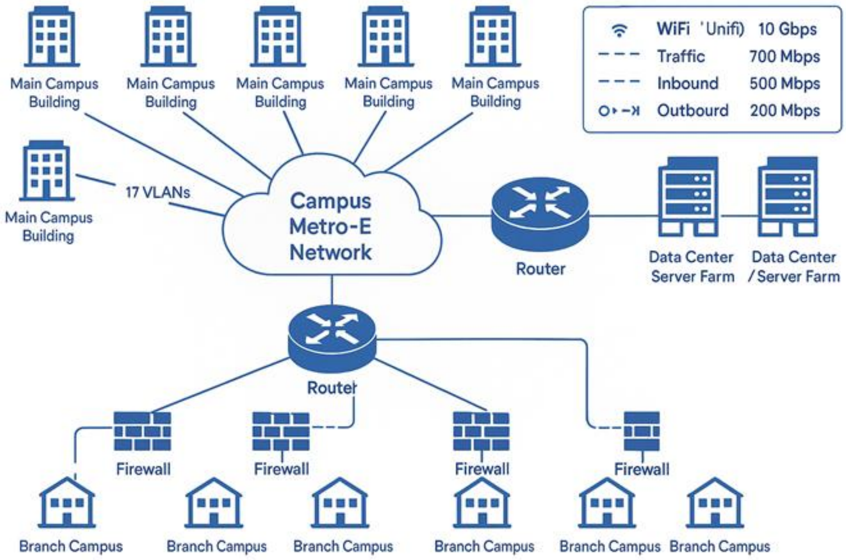

Figure 2. Illustration of Metro-E network campus

Table 2. Summary of variable parameter

Parameter	Symbol	Values
Time-based traffic	X	0:00 <X <11.59 pm
Number of days	D	90
Inter-arrival time	Ta	5 minutes
Time frame minimum	Tmin	0:00 am
Time frame maximum	Tmax	11.50 pm
number of data minimum	α	2.33 and 1.9909
number of data maximum	β	1058.523 and 379.8954
Policing inbound threshold campus A	ThtA	500 Mbps
Policing outbound threshold campus B-H	ThtX	200 Mbps
Access rate campus A	Ca	10 Gbps
Access rate campus B-H	Cx	2 Gbps

3. RESULTS AND DISCUSSION

3.1. Inbound and outbound data correlation analysis

Figure 3 shows that the variables for inbound and outbound data are visualized using color shades that indicate the strength and direction of their correlations. In analyzing inbound data, several summary statistics reveal relationships between variables across locations A through H. The average correlation values are as follows: A (0.4967), B (0.2442), C (0.2408), D (0.3971), E (0.4958), F (0.4547), G (0.2314), and H (0.5204), with location H showing the highest mean correlation. This suggests H significantly influences other variables. Conversely, location F has the highest standard deviation at 0.4076, indicating more volatility in its correlations. Locations G and C display the lowest correlation at 0.0240, implying a negligible relationship between them. Figure 4 presents the analysis of outbound data reveals an understanding of the relationships among locations, with mean correlations ranging from 0.1537 to 0.4757. Location A exhibits the highest mean correlation at 0.4757, while location C has the lowest at 0.1537, indicating moderate interconnectedness, though slightly less than inbound data. Location F shows the highest standard deviation of 0.4395, highlighting varying correlations across locations.

Weak negative correlations exist between pairs B and E, and C and F, but they do not significantly impact overall trends. Understanding internet traffic patterns is crucial for optimizing network performance. Locations A, E, F, and H are major hubs with strong connectivity, while G, C, and B show weak correlations, indicating they function as isolated nodes or backup servers. Inbound traffic correlates strongly among A, E, F, and H, reflecting high user engagement, while G shows weak inbound traffic. Outbound traffic is more complex, featuring lower and negative correlations, which may indicate issues like traffic congestion.

Table 3 summarizes the means, correlation, and standard deviation of inbound and outbound data for all campuses in the Metro-E network. The analysis reveals strong correlations in locations A, E, F, and H, with H serving as a core hub. In contrast, B, C, and G show weak correlations, indicating isolated roles. Location F has the highest standard deviation (0.4076), highlighting unpredictability in its traffic. Negative correlation pairs, like B and E, and C and F, suggest potential congestion issues, indicating a need for strategic traffic management. Overall, these findings illustrate the complexity of traffic dynamics and varying interdependencies among locations.

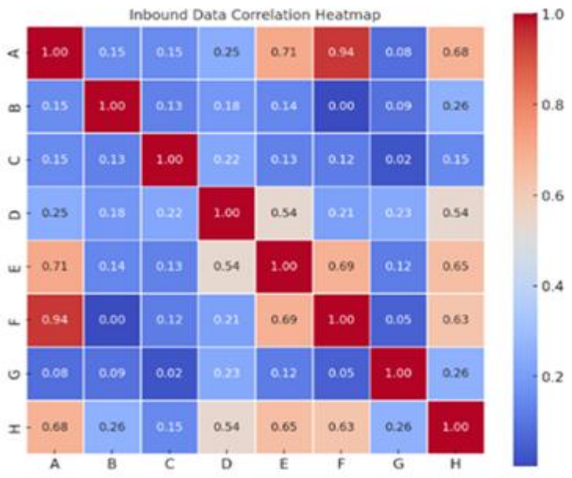

Figure 3. Correlation analysis of inbound data

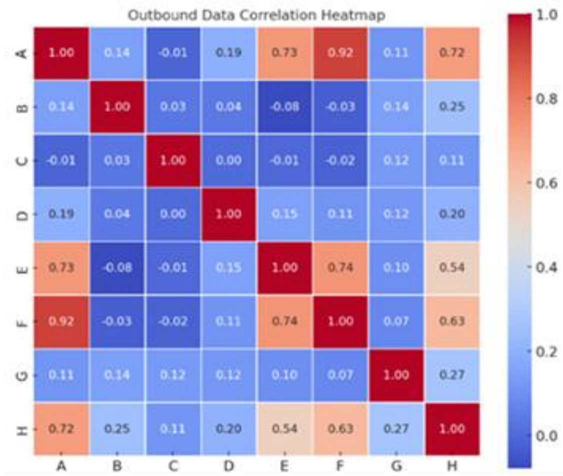

Figure 4. Correlation analysis of outbound data

Table 3. Correlation analysis of inbound and outbound data traffic

Location	Mean correlation	Standard deviation	Key observations
A	0.4967	0.3021	Strong correlation, high traffic activity.
B	0.2442	0.1984	Weak correlation, independent fluctuations.
C	0.2408	0.1950	Weak correlation, contains outliers.
D	0.3971	0.2783	Moderate correlation, some irregular spikes.
E	0.4958	0.3015	Strong correlation, stable traffic behavior.
F	0.4547	0.4076	High variability, frequent fluctuations.
G	0.2314	0.0240	Very weak correlation, sporadic traffic usage.
H	0.5204	0.3127	Strongest correlation acts as a core network hub.

3.2. Descriptive analysis of inbound and outbound data for all locations on specified dates

Figures 5 to 9 analyze inbound and outbound Internet traffic over specific dates, highlighting fluctuations in volume and important trends. These visualizations clarify traffic behavior, aiding stakeholders in making informed decisions about resource allocation and network management. Figure 10 shows stable internet traffic with minor anomalies, indicating a consistent flow. This analysis is crucial for effective traffic management. In contrast, Figures 11 and 12 reveal skewed traffic distribution, highlighting congestion at two locations, leading to potential delays during peak times.

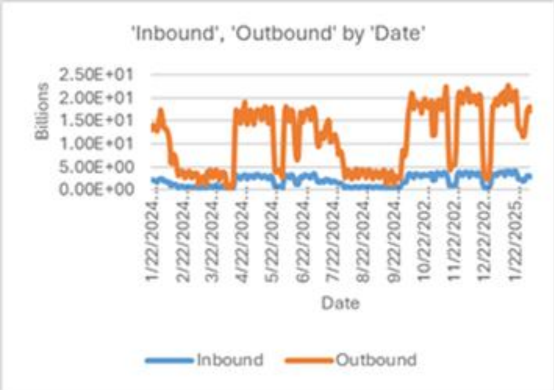

Figure 5. Campus A in/out data traffic

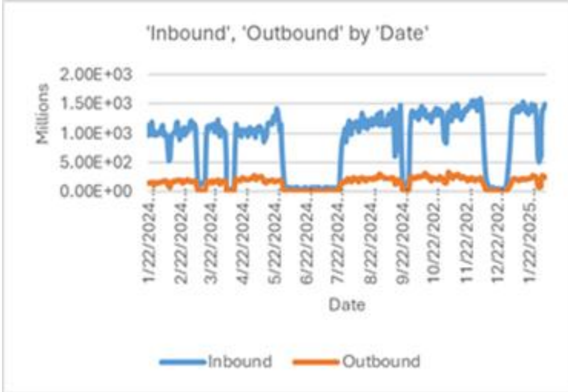

Figure 6. Campus B in/out data traffic

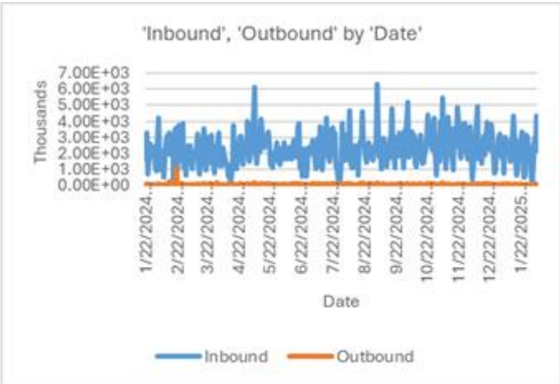

Figure 7. Campus C in/out data traffic

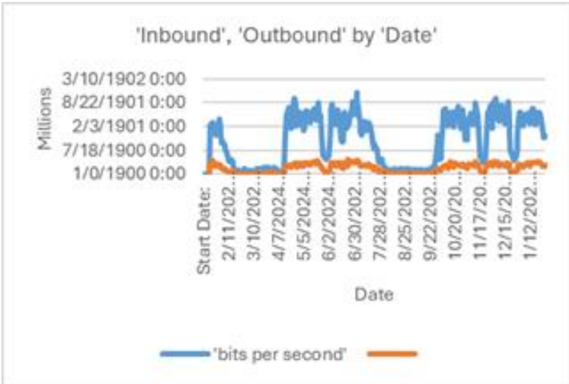

Figure 8. Campus D in/out data traffic

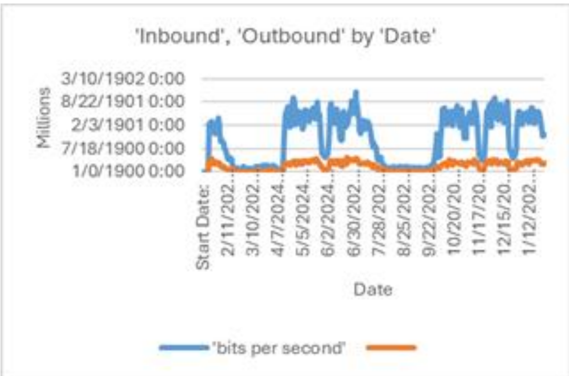

Figure 9. Campus F in/out data traffic

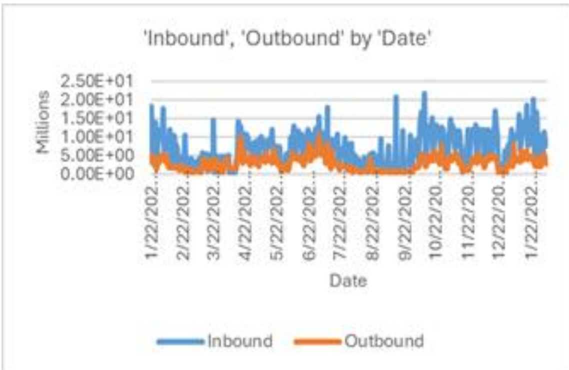

Figure 10. Campus E in/out data traffic

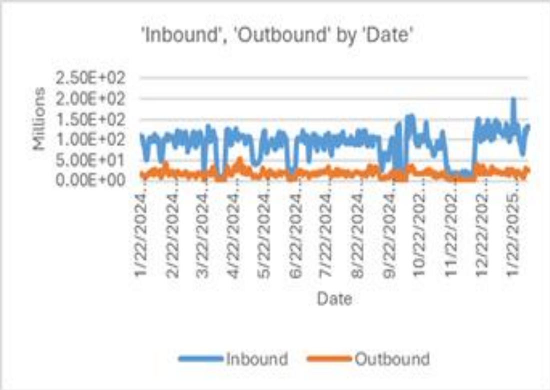

Figure 11. Campus G in/out data traffic

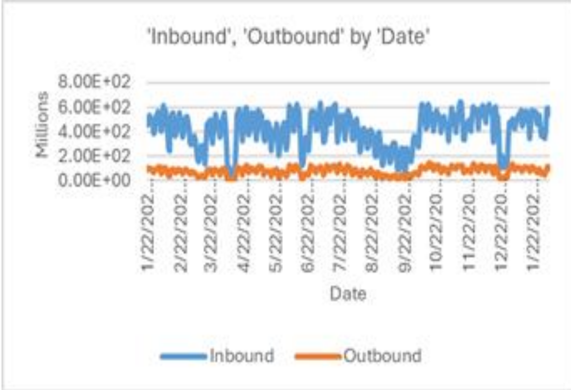

Figure 12. Campus H in/out data traffic

Figures 13 to 18 reveal a strong correlation between inbound and outbound internet traffic within the analyzed date range. This relationship is vital for understanding traffic patterns and enhancing network performance. Campus A exhibits a high correlation coefficient of 0.987, while Campuses B and H show moderate correlations of 0.954 and 0.932, respectively. Campus E's correlation is lower at 0.662, and Campus D has the weakest correlation at 0.348, indicating a weaker connection between its incoming and outgoing traffic patterns. Figures 19 and 20 show a strong correlation between inbound and outbound internet data traffic, but outliers indicate potential network issues. Analyzing these anomalies, such as network failures and unusual user behavior, can improve network optimization.

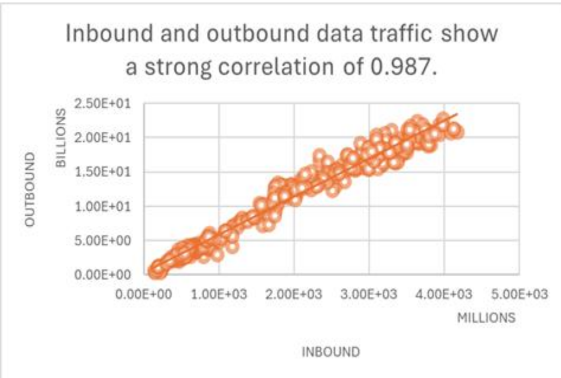

Figure 13. Correlation analysis on campus A

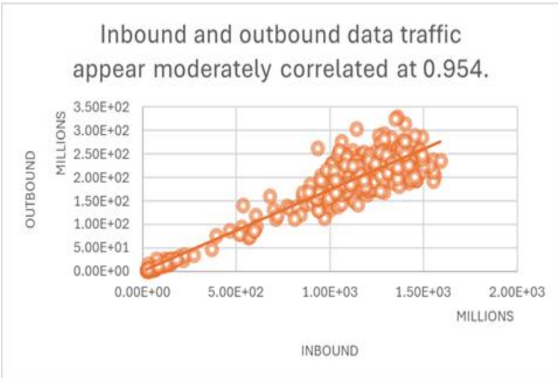

Figure 14. Correlation analysis on campus B

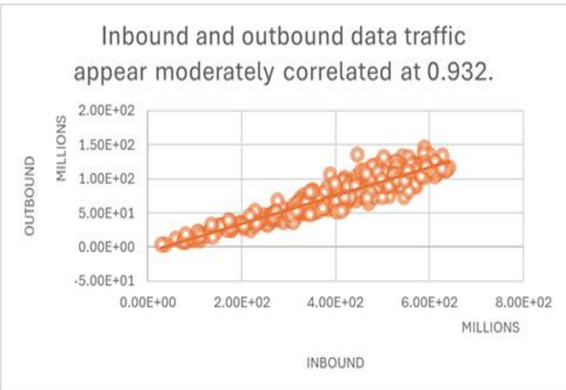

Figure 15. Correlation analysis on campus H

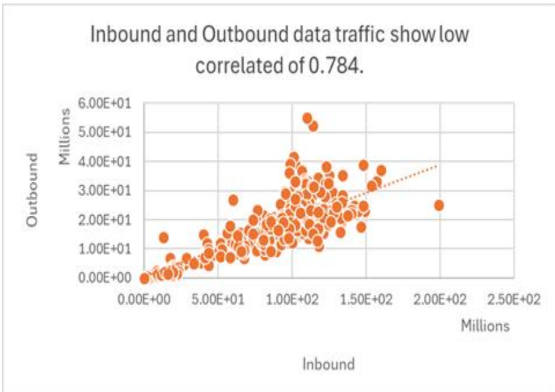

Figure 16. Correlation analysis on campus G

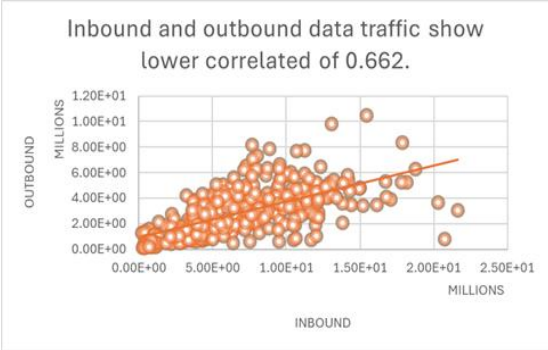

Figure 17. Correlation analysis on campus E

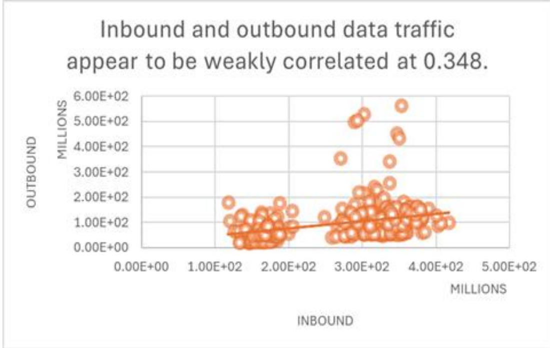

Figure 18. Correlation analysis on campus D

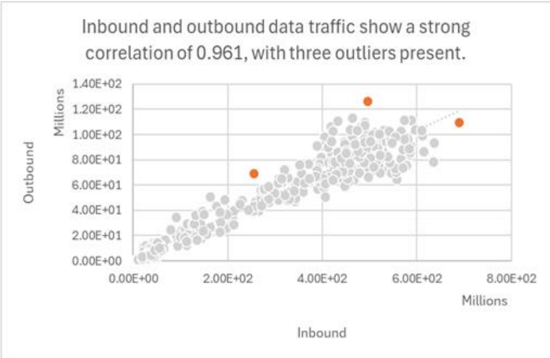

Figure 19. Campus C in/out data highly correlated

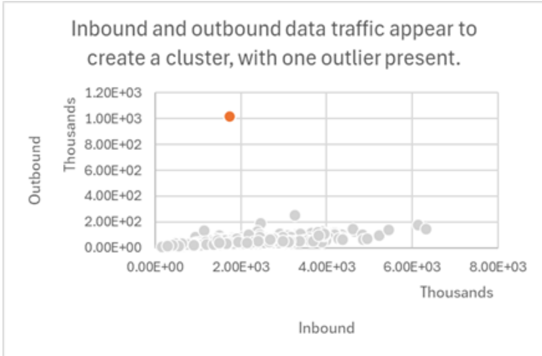

Figure 20. Campus F form a cluster with 1 outlier

3.3. Possible network behaviors and issues

The analysis reveals critical insights for network management. Locations A, B, C, F, and H have strong correlations that may lead to traffic congestion, necessitating efficient traffic management. In contrast, locations D, E, and G show weak correlations, indicating underutilization, potentially causing bottlenecks. Location F requires strategic load-balancing to manage fluctuating traffic. ML techniques like ARIMA, LSTM, and CNN can analyze traffic patterns; while ARIMA offers moderate short-term forecasts, LSTM and CNN are more effective for identifying patterns. Seasonal traffic variations can impact forecast accuracy. To enhance network performance, it's advisable to implement dynamic bandwidth allocation, use real-time monitoring for anomaly detection, and improve load balancing in high-traffic areas.

4. CONCLUSION

This study highlights the effectiveness of hybrid ML models, such as ARIMA-LSTM and ARIMA-CNN, for internet traffic forecasting in Metro-E campus networks. By combining statistical time-series methods with deep learning techniques, these models overcome the limitations of traditional approaches in handling complex traffic patterns. Analysis of data from eight campus locations showed significant variations, with strong correlations in Locations A, E, F, and H, while B, C, and G exhibited more independent behaviors. Outliers in C and F pointed to potential network anomalies. Notably, Location H had the highest average traffic (75 Mbps), and Location F showed the most variability. The improved forecasting accuracy of hybrid models allows for better bandwidth management and user experience. However, challenges such as the need for efficient computational resources and real-time adaptations remain. Future research should focus on enhancing model scalability and integrating external factors for improved traffic forecasting and bandwidth allocation, ultimately optimizing network performance in educational settings.

ACKNOWLEDGMENTS

Authors acknowledge the Institut Pengajian Siswazah (IPSIS), Universiti Teknologi MARA (UiTM) for the Journal Support Fund (JSF) in funding this publication.

FUNDING INFORMATION

Authors state there is no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Norakmar Arbain	✓	✓	✓	✓	✓	✓		✓	✓	✓				
Murizah Kassim	✓	✓		✓			✓	✓		✓	✓	✓	✓	✓
Darmawaty Mohd Ali		✓		✓						✓		✓		
Shuria Saaidin	✓	✓		✓				✓		✓		✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

INFORMED CONSENT

We have obtained informed consent from all individuals included in this study.

ETHICAL APPROVAL

There are no ethical issues related to this research.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [MK], upon reasonable request.

REFERENCES

[1] G. O. Ferreira *et al.*, "Forecasting network traffic: a survey and tutorial with open-source comparative evaluation," *IEEE Access*, vol. 170, no. November 2022, pp. 19–41, 2023, doi: 10.1016/j.comcom.2021.01.021.

[2] N. P. Abdullah, S. M. Deni, and M. Kassim, "WAN internet traffic parameter analysis on Metro-E campus network," in *2023 IEEE 14th Control and System Graduate Research Colloquium (ICSGRC)*, 2023, pp. 180–185. doi: 10.1109/ICSGRC57744.2023.10215487.

[3] R. Alvarado and A. Suárez, "A novel energy-saving method for campus wired and dense WiFi network applying machine learning and idle cycling techniques," *FACETS*, vol. 9, pp. 1–19, 2024, doi: 10.1139/facets-2023-0164.

[4] H. Mliki, L. Chaari, and L. Kamoun, "A comprehensive survey on carrier ethernet congestion management mechanism," *Journal of Network and Computer Applications*, vol. 47, pp. 107–130, 2015, doi: 10.1016/j.jnca.2014.09.003.

[5] J. Zheng and M. Huang, "Traffic flow forecast through time series analysis based on deep learning," *IEEE Access*, vol. 8, pp. 82562–82570, 2020, doi: 10.1109/ACCESS.2020.2990738.

[6] S. T. Aung and T. Thein, "Internet traffic categories demand prediction to support dynamic QoS," in *2020 5th International Conference on Computer and Communication Systems (ICCCS)*, 2020, pp. 650–654. doi: 10.1109/ICCCS49078.2020.9118431.

[7] N. P. Abdullah, M. Kassim, S. M. Deni, and Y. M. Yussoff, "Descriptive analysis of wide area network flow control internet traffic on Metro-E 100 Mbps campus network," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 4, pp. 2738–2749, 2024, doi: 10.11591/eei.v13i4.7044.

[8] A. D. S. A. Shamsuddin, M. Kassim, and N. P. Abdullah, "Bandwidth performance analysis and shaping algorithm on Metro-E campus network," in *2023 IEEE Congress on Information Science and Technology (CiSt)*, 2023, pp. 7–12. doi: 10.1109/CiSt56084.2023.10409961.

[9] R. A. Rahman, M. Kassim, Y. C. K. H. C. Ku, and M. Ismail, "Performance analysis of routing protocol in WiMAX network," in *2011 IEEE International Conference on System Engineering and Technology (ICSET)*, 2011, pp. 153–157. doi: 10.1109/ICSEngT.2011.5993440.

[10] N. A. Salim, V. N. Sulistyawan, and F. T. Intan, "Analysis and solutions of traffic shift on 4G networks in the campus environment during the COVID-19 pandemic," *IOP Conference Series: Earth and Environmental Science*, vol. 969, no. 1, p. 12027, 2022, doi: 10.1088/1755-1315/969/1/012027.

[11] S. Ahmad, B. Scotney, D. Glass, and S. Zhang, "Enhancing network performance monitoring through scalable multi-dimensional metric analysis and pattern-based anomaly detection." 2024. doi: 10.21203/rs.3.rs-4914517/v1.

[12] Y. Geng and S. Li, "A LSTM based campus network traffic prediction system," in *2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS)*, 2019, pp. 327–330. doi: 10.1109/ICSESS47205.2019.9040735.

[13] K. P. N. Rao, S. A. Kumar, and K. Keerthika, "Adaptive QoS protocol in systemized networks with machine learning," in *2024 Second International Conference on Networks, Multimedia and Information Technology (NMITCON)*, 2024, pp. 1–6. doi: 10.1109/NMITCON62075.2024.10698888.

[14] M. Kushwaha, A. Sangwan, and K. Gupta, "Real time web traffic forecasting using LSTM and CNN," in *2025 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, 2025, pp. 1–7. doi: 10.1109/SCEECS64059.2025.10940845.

[15] H. G. A. Elrahim, N. N. N. A. Malik, and K. B. M. Yusof, "Efficient network traffic prediction: Harnessing hybrid neural networks," in *2024 IEEE International Conference on Advanced Telecommunication and Networking Technologies (ATNT)*, 2024.

[16] A. R. Sattarzadeh, R. J. Kutadinata, P. N. Pathirana, and V. T. Huynh, "A novel hybrid deep learning model with ARIMA Conv-LSTM networks and shuffle attention layer for short-term traffic flow prediction," *Transportmetrica A: Transport Science*, vol. 21, no. 1, p. 2236724, 2025, doi: 10.1080/23249935.2023.2236724.

[17] J. Wu, T. Qiu, H. Tang, and X. Liu, "Network traffic prediction based on a CNN-LSTM with attention mechanism," in *2022 7th International Conference on Computational Intelligence and Applications (ICCIA)*, 2022.

[18] S. Saha and A. Haque, "Wavelet-based hybrid machine learning model for out-of-distribution internet traffic prediction," in *Proceedings of IEEE/IFIP Network Operations and Management Symposium 2023 (NOMS)*, 2023, pp. 1–8. doi: 10.1109/NOMS56928.2023.10154337.

[19] M. G. A. Pazou *et al.*, "Network traffic prediction by learning time series as images," *Engineering Science and Technology, an International Journal*, vol. 55, no. February, p. 101754, 2024, doi: 10.1016/j.jestch.2024.101754.

[20] N. M. Balamurugan, M. Adimoolam, M. H. Alsharif, and P. Uthansakul, "A novel method for improved network traffic prediction using enhanced deep reinforcement learning algorithm," *Sensors*, vol. 22, no. 13, 2022, doi: 10.3390/s22135006.

[21] T. H. H. Aldhyani, M. Alrasheedi, A. A. Alqarni, M. Y. Alzahrani, and A. M. Bamhdi, "Intelligent hybrid model to enhance time series models for predicting network traffic," *IEEE Access*, vol. 8, pp. 130431–130451, 2020, doi: 10.1109/ACCESS.2020.3009169.

[22] K. Albeladi, B. Zafar, and A. Mueen, "Time series forecasting using LSTM and ARIMA," *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 1, pp. 313–320, 2023, doi: 10.14569/IJACSA.2023.0140133.

[23] W. Wang *et al.*, "A network traffic flow prediction with deep learning approach for large-scale metropolitan area network," in *IEEE/IFIP Network Operations and Management Symposium: Cognitive Management in a Cyber World (NOMS)*, 2018, pp. 1–9. doi: 10.1109/NOMS.2018.8406252.

[24] M. Kassim, M. Ismail, and M. I. Yusof, "Statistical analysis and modeling of internet traffic IP-based network for tele-traffic engineering," *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 3, pp. 1505–1512, 2015.

[25] M. Usama *et al.*, "Unsupervised machine learning for networking: techniques, applications and research challenges," *IEEE Access*, vol. 7, pp. 65579–65615, 2019, doi: 10.1109/ACCESS.2019.2916648.

[26] M. Abbasi, A. Shahraki, and A. Taherkordi, "Deep learning for network traffic monitoring and analysis (NTMA): a survey," *Computer Communications*, vol. 170, no. December 2020, pp. 19–41, 2021, doi: 10.1016/j.comcom.2021.01.021.

[27] Y. Xu, "Machine learning based traffic prediction and congestion control algorithms in software defined networks," in *2024 IEEE International Conference on Information and Intelligent Systems Technologies (IIST)*, 2024, pp. 285–289. doi: 10.1109/IIST62526.2024.00035.

[28] H. Yang, X. Li, W. Qiang, Y. Zhao, W. Zhang, and C. Tang, "A network traffic forecasting method based on SA optimized ARIMA–BP neural network," *Computer Networks*, vol. 193, no. February, p. 108102, 2021, doi: 10.1016/j.comnet.2021.108102.

[29] S. Khorsandroo, A. G. Sánchez, A. S. Tosun, J. M. Arco, and R. D.-Corin, "Hybrid SDN evolution: a comprehensive survey of the state-of-the-art," *Computer Networks*, vol. 192, p. 107981, 2021, doi: 10.1016/j.comnet.2021.107981.

[30] D. Aloraifan, I. Ahmad, and E. Alrashed, "Deep learning based network traffic matrix prediction," *International Journal of Intelligent Networks*, vol. 2, no. May, pp. 46–56, 2021, doi: 10.1016/j.ijin.2021.06.002.

[31] S. Saha, A. Haque, and G. Sidebottom, "Transfer learning based efficient traffic prediction with limited training data," in *Proceedings of IEEE Consumer Communications and Networking Conference (CCNC)*, 2023, vol. 2023-Janua, pp. 477–480. doi: 10.1109/CCNC51644.2023.10060745.

[32] S. Saha, A. Haque, and G. Sidebottom, "Analyzing the impact of outlier data points on multi-step internet traffic prediction using deep sequence models," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1345–1362, 2023, doi: 10.1109/TNSM.2023.3262406.

[33] L. Wang, L. Che, K. Y. Lam, W. Liu, and F. Li, "Mobile traffic prediction with attention-based hybrid deep learning," *Physical Communication*, vol. 66, no. February, p. 102420, 2024, doi: 10.1016/j.phycom.2024.102420.

[34] S. Fischer, K. Katsarou, and O. Holschke, "DeepFlow: towards network-wide ingress traffic prediction using machine learning at large scale," in *2020 International Symposium on Networks, Computers and Communications (ISNCC)*, 2020. doi: 10.1109/ISNCC49221.2020.9297301.

[35] A. Azari, P. Papapetrou, S. Denic, and G. Peters, "Cellular traffic prediction and classification: a comparative evaluation of LSTM and ARIMA," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11828 LNAI, 2019, pp. 129–144. doi: 10.1007/978-3-030-33778-0_11.

[36] Y. Miao *et al.*, "A novel short-term traffic prediction model based on SVD and ARIMA with blockchain in industrial internet of things," *IEEE Internet of Things Journal*, vol. 10, no. 24, pp. 21217–21226, 2023, doi: 10.1109/JIOT.2023.3283611.

[37] C. Hajaj, P. Aharon, R. Dubin, and A. Dvir, "The art of time-bending: Data augmentation and early prediction for efficient traffic classification," *Expert Systems with Applications*, vol. 252, no. April, 2024, doi: 10.1016/j.eswa.2024.124166.

[38] J. Shi, Y. B. Leau, K. Li, and J. H. Obit, "A comprehensive review on hybrid network traffic prediction model," *International Journal of Electrical and Computer Engineering*, vol. 11, no. 2, pp. 1450–1459, 2021, doi: 10.11591/ijece.v11i2.pp1450-1459.

[39] S. Saha, S. Das, and G. H. Carvalho, "ConvLSTMTransNet: a hybrid deep learning approach for internet traffic telemetry," in *2024 IEEE Virtual Conference on Communications (VCC)*, 2024, pp. 1–6.

[40] B. Shao, D. Song, G. Bian, and Y. Zhao, "A hybrid approach by CEEMDAN-improved PSO-LSTM model for network traffic prediction," *Security and Communication Networks*, vol. 2022, no. 1, p. 4975288, 2022, doi: 10.1155/2022/4975288.

[41] J. Su, H. Cai, Z. Sheng, A. X. Liu, and A. Baz, "Traffic prediction for 5G: a deep learning approach based on lightweight hybrid attention networks," *Digital Signal Processing*, vol. 146, p. 104359, 2024, doi: 10.1016/j.dsp.2023.104359.

BIOGRAPHIES OF AUTHORS

Norakmar Arbain is a Lecturer at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. She received her M.Sc. in Real-Time Software Engineering from Universiti Teknologi Malaysia (UTM). She is currently pursuing her Ph.D. degree in the field of Computer Engineering at Universiti Teknologi MARA (UiTM). Her research interests cover software engineering, computer programming, mobile networking, and machine learning applications. Recently, she received the best presenter award at the Soft Computing in Data Science (SCDS2025) Conference. She can be contacted at email: norakmar7222@uitm.edu.my.

Murizah Kassim currently working as Head of Publication and Innovation and Senior Fellow at the Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Universiti Teknologi MARA, Shah Alam. She is an Associate Professor from the Faculty of Electrical Engineering, Universiti Teknologi MARA, Shah Alam, Selangor. She received her Ph.D. in Electronic, Electrical, and Systems Engineering from the Faculty of Built Environment and Engineering, Universiti Kebangsaan Malaysia (UKM). Her research areas are computer networks, data analytics, IoT, web, and mobile development applications research. She has won many awards for Best Paper, exhibitions, industries, and grant awards throughout her career at UiTM. She is also a member of the Enabling Internet of Things Technologies (EIIoTT) research group at UiTM. She is currently a member of MBOT, IEEE, IET, IAENG, and IACSIT organizations. She can be contacted at e-mail: murizah@uitm.edu.my.

Darmawaty Mohd Ali is an Associate Professor at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM). In 2012, she was awarded a Ph.D. degree from Universiti Malaya (UM). She received her first degree from Universiti Kebangsaan Malaysia (UKM), with Honors, in Electrical, Electronic, and Systems, graduating in 1999. She started her first job as a Product Engineer before furthering her studies for her master's degree at the University Teknologi Malaysia (UTM). Her research interest includes wireless access technology and QoS provisioning in wireless networks. She can be contacted at email: darma504@uitm.edu.my.

Shuria Saaidin is a Lecturer at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, Malaysia. She holds a Ph.D. in Computer Science from UiTM and has over ten years of experience in teaching and research. Her research interests include cloud computing, artificial intelligence, machine learning, data analytics, and the internet of things (IoT). She has published in international journals and IEEE conferences and actively contributes to educational innovation and applied research in computing technologies. She can be contacted at email: shuria6809@uitm.edu.my.