

Analysis of hybrid non-linear autoregressive neural network and local smoothing technique for bandwidth slice forecast

Mohamed Khalafalla Hassan¹, Sharifah H. S. Ariffin², Sharifah Kamilah Syed-Yusof³,
N. Effiyana Ghazali⁴, Mohammed EA Kanona⁵

^{1,2,3,4}School of Electrical Engineering, University Technology Malaysia, Malaysia

^{1,5}Faculty of Telecommunication Engineering, Future University, Khartoum, Sudan

Article Info

Article history:

Received Jun 20, 2020

Revised Apr 7, 2021

Accepted Apr 21, 2021

Keywords:

Autoregressive neural network

Bandwidth slice

Forecast

Local smoothing

ABSTRACT

The demand for high steady state network traffic utilization is growing exponentially. Therefore, traffic forecasting has become essential for powering greedy application and services such as the internet of things (IoT) and Big data for 5G networks for better resource planning, allocation, and optimization. The accuracy of forecasting modeling has become crucial for fundamental network operations such as routing management, congestion management, and to guarantee quality of service overall. In this paper, a hybrid network forecast model was analyzed; the model combines a non-linear auto regressive neural network (NARNN) and various smoothing techniques, namely, local regression (LOESS), moving average, locally weighted scatterplot smoothing (LOWESS), the Sgolay filter, Robyn loess (RLOESS), and robust locally weighted scatterplot smoothing (RLOWESS). The effects of applying smoothing techniques with varied smoothing windows were shown and the performance of the hybrid NARNN and smoothing techniques discussed. The results show that the hybrid model can effectively be used to enhance forecasting performance in terms of forecasting accuracy, with the assistance of the smoothing techniques, which minimized data losses. In this work, root mean square error (RMSE) is used as performance measures and the results were verified via statistical significance tests.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Mohamed Khalafalla Hassan
School of Electrical engineering
University Technology Malaysia
Johor Bahru, Malaysia
Email: memo1023@hotmail.com

1. INTRODUCTION

Nowadays, there are vast network deployments of various domains and emerging new technologies and application-centric services. The capability of network traffic forecast has become one of today's crucial network design and the main requirements of various operations due to its benefits in various sub-domains, such as network security, dynamic slice re-allocation, and resource planning. In network traffic forecast, a proactive approach is used instead of a reactive one, where network resources are monitored to ensure that all service requirements are met, in addition to quality of service (QoS) and security. Also, traffic analysis can be a crucial stage for building successful preventive congestion controls.

Generally, forecast and prediction are used interchangeably. Nevertheless, forecast can be explicitly defined as the estimation of future values based on an analytical model built from past observations. In this

paper, a forecast approach based on past values was adopted. Generally, there are two broad approaches to forecast network traffic: long-period forecast and short-period forecast. Long-period traffic forecasts can be used to assess future resource demands, and accordingly allows for additional planning time and hence better decisions. On the other hand, short-period forecasts can be associated with dynamic resource re-allocation. Moreover, this type of forecast is normally used to enhance quality of service (QoS), improve congestion control mechanisms, and optimize network resource management and routing decision management.

Various approaches have been used for network traffic analysis and forecasting, such as time series models, modern data mining techniques, machine learning (ML), and hybrid techniques.

The main focus of this paper is on hybrid machine learning-based frameworks. Generally, ML techniques are used in various domains to solve various complex problems including optimization resource management, allocation, and automation. ML applications are also used in communication networks.

The application of ML has recorded an unprecedented surge in communication networks. ML enables a system to summarize and abstract data to deduce knowledge [1]. It also provides the researcher with the ability to improve knowledge over time and with experience, with the objective of discovering hidden patterns and exploring unknown data. Therefore, ML is gaining more attention in areas involving data analysis, fitting, decision-making, and automation.

Machine learning is expected have a more dominant role in future emerging telecommunication technologies and architectures such as in internet of things (IoT), block chain, and 5G network operations and management [1]. In today's complex network architecture and emerging demands for various services, network traffic forecast has become increasingly vital to ensure smooth network operations and management. Generally, forecasting is seen as a time series data and is accordingly modeled via time series forecast techniques to establish a correlation between previously observed traffic and future demands.

Time series analysis is still considered a challenge because it involves complicated combinations of nonlinear and non-stationary dynamic behaviors. Statistically, dynamic systems produce a non-linear time series if the output is characterized by non-linear features such as non-normality, aperiodicity, and nonlinear causal relationships between lagged variables.

Generally, two broad approaches have been used for developing statistical analysis models and supervised ML models. Statistical analysis models are based on the generalized autoregressive integrated moving average (ARIMA) model, while the majority of traffic forecasting models are based on supervised ML and more specifically on artificial neural networks (ANNs). However, ARIMA-based models fall short when dealing with nonlinear and non-stationary data [1], [2]. The main difference between neural network autoregressive (NNAR) and autoregressive integrated moving average (ARIMA) is that the former requires a stationary property to be imposed. Historically, different types of ANNs and other ML techniques have been used for forecasting the time series of network traffic.

Preprocessing has become crucial in data science, signal processing, and machine learnin due to incomplete, inconsistent (containing errors, outlier values), and varying noise patterns that exist and are embedded in collected data. Hence, preprocessing methods must be employed before network forecasting can be done to enhance data quality. In turn, this step will enhance the accuracy and efficiency of non-linear autoregressive neural network (NARNN). Preprocessing techniques are considered crucial.

The scope of this paper is limited to network bandwidth forecast and not the general problem of time series forecast. Therefore, the distinctive features and limitations of notable previous research were summarized. The data used was collected from a premier internet service provider representing an long term evolution (LTE) 4G core aggregated bandwidth slice. Two forecast time scales were used: one day and one week. The collected data was used to develop a forecast model, namely a univariate time series model. A hybrid nonlinear autoregressive neural network was used for forecasting combined with various dynamic smoothing techniques. Smoothing techniques were used to enhance forecast accuracy.

2. RELATED WORK

Cortez *et al.* [3] used multi layer perceptron neural network (MLP-NN) and simple network management protocol (SNMP) traffic gathered from two different internet service provider (ISP) networks as a dataset. Two subsets were investigated-one subset representing traffic on a trans-Atlantic link, and another representing aggregated traffic in the backbone of the ISP. Missing SNMP was completed using linear interpolation. The performance of the proposed model was compared to traditional Holt-Winters models, and double Holt-Winters ARIMA models. The results showed that the NN model outperformed traditional ARMA models. However, the proposed model was static and did not react to the dynamic nature of traffic loads.

Chabaa *et al.* [4] evaluated various back propagation (BP) training algorithms MLP-NN for an Internet traffic time series. The proposed work showed that Levenberg-Marquardt (LM) and resilient propagation (Rp) algorithms outperformed other BP algorithms. Zhu *et al.* [5], a hybrid training algorithm was proposed based

on an artificial bee colony (ABC) algorithm that employed particle swarm optimization (PSO), an evolutionary search algorithm. Moreover, a (5;11;1) MLP-NN was used as the training algorithm. The results showed that the proposed model had a higher prediction accuracy than BP.

Li *et al.* [6] used a feed forward neural network to predict incoming and outgoing traffic flows. The study argued that inter-data center link is dominated by elephant flows. The study used a gradient decent and a wavelet transform to train a hybrid model. SNMP counters and total incoming and outgoing data traffic were gathered in 30-second intervals. These data were used as the dataset. The data were collected from data center (DC) routers for a period of six weeks. The time series was decomposed using a level-10 wavelet transform. However, it must be noted that the wavelet transform can aggressively eliminate parts of the original data if not implemented carefully.

Dyllon *et al.* [7] developed a nonlinear autoregressive exogenous neural (NARX) network model for time series network traffic analysis. The study implemented a neural network model to predict the future trends of the London South Bank University (LSBU) bandwidth data traffic. Dataset was collected using the paessler router traffic grapher (PRTG) tool. The results showed that NARX neural network is a good method for predicting time series data.

Yoo and Sim [8] proposed a forecast model and claimed it could improve resource utilization efficiency in high-bandwidth networks to accommodate the rise in data volume demands for scientific data applications. A seasonal decomposition of time series by LOESS (STL) and ARIMA are used on SNMP. The results showed that the proposed forecast model was resilient against abrupt changes in network usage. The multistep forecast was tested as well.

Afolabi *et al.* [9] discussed the significance of the interference-less machine learning approach in a time series forecast as a crucial component of prediction performance, especially when forecasting many steps ahead of the currently available data. The authors used Hilbert Huang transformation (HHT) as the noise elimination technique. The simulation results were compared with conventional and state-of-the-art approaches.

Joo *et al.* [10] proposed a prediction method based on wavelet filtering. The proposed framework analyzed the time series in both the time and frequency domains. The proposed approach was applied to various scenarios. The results showed that the proposed method outperformed other approaches that did not use wavelet-filtering techniques. B. Doucoure *et al.* [11] introduced a prediction method for renewable energy sources to intelligently manage renewable energy. The authors used wavelet decomposition and artificial neural networks and discussed the significance of their results.

Alawe *et al.* [12] proposed a novel mechanism to scale 5G core network resources by forecasting traffic via ML techniques. The prediction technique used was based on recurrent neural networks (RNN), long short-term memory (LSTM), artificial neural networks (ANN), and the deep neural network (DNN). Comparisons were made between the different techniques. The simulation results confirmed the higher efficiency of the RNN-based solution compared to the other approaches. No preprocessing or feature extraction was made.

Wang *et al.* [13] proposed a wavelet-based neural network model, called the multilevel wavelet decomposition network (mWDN). The proposed model used the wavelet decomposition in frequency learning while enabling the fine-tuning of all parameters under a deep neural network framework. The results showed the effectiveness of the proposed hybrid approach. The wavelet decomposition required several parameters that could affect the forecast performance such as the number of decomposition levels and the selected mother wavelet.

Salih [14] introduced LAN office network bandwidth prediction models as time series models. The proposed forecast models were tested using mean square error (MSE) and performance evaluation plots. However, the study did not use any preprocessing techniques. J. Feng *et al.* [15] proposed a deep traffic predictor (DeepTP) model to forecast long-period cellular network traffic. The study showed that the model outperformed other traffic forecast models by more than 12.3%. However, LSTM is not suitable for long-period forecasting (multi-steps ahead).

Le *et al.* [16] proposed a traffic forecasting model using autoregressive models and neural network, models to predict key performance indicators (KPIs) in network KPI for long term and short-term forecasting real data. However, no preprocessing was applied and the study only focused on investigating relationships between network KPIs.

You *et al.* [17] proposed a hybrid LOESS-ARIMA-based forecast model. Authors claimed that such a model has the potential of enhancing the efficiency of resource utilization, especially in high-speed networks, to accommodate the rapid increase in rising demands for scientific data applications. A seasonal decomposition of time series by LOESS (STL) and (ARIMA) was applied on simple network management protocol (SNMP). The results revealed that the proposed forecast model was resilient against abrupt changes in network usage provided that the multistep forecast was used as the primary scenario.

3. RESEARCH METHOD

The ML approaches were modeled as time series batch learning. The general process of network bandwidth forecast is based on machine learning. This algorithm was extended in this study by preprocessing the provided dataset, namely by eliminating unnecessary noise and rapid traffic fluctuations. Moreover, to avoid the erosion of periodic trends and patterns within the series, the system learns local and global trends separately to detect and eliminate short-term or long-term noise. Similar approaches have also been used in the past [7]-[11], with the various techniques used including Hilbert Huang transformation (HHT), STL, and the wavelet-based approach. However, it is often used to detect high noise levels in the long term and may not be suitable for online or semi-online processes, while the current study proposes a hybrid approach using a nonlinear auto regressive neural network that focuses mainly on local variations using various local regression techniques to remove unnecessary noise and fluctuations, which may have negative effects on the prediction accuracy, especially in nonlinear and non-stationary time series. Local regression approaches allow the removal of noise and fluctuations in short scales and react more dynamically to noise-level short-term variations more than other wavelet- and HHT-based techniques. Similar approaches were also utilized in one study [8], which used ARIMA instead of NAR. The effectiveness of the proposed method was verified using available real network traffic datasets.

3.1. Neural network auto-regressive (NNAR)

Neural network training attempts to approximate a function by optimizing network weights and neuron bias.

$$y(t) = f(y(t-1), \dots, y(t-d)) + \varepsilon \quad (1)$$

In (1), the term ε stands for error. The y input features (Bandwidth slice in this case) $y(t-1)$, $y(t-2)$, $y(t-3)$ are the feedback delays. Trial-and-error was done to optimize the hidden layers and neurons to achieve the best performance. However, as the number of neurons increases as the system becomes more complex, the low number of neurons may reduce network efficiency. Levenberg-Marquardt is the most widely used learning rule due to its fast response [9]. The root mean squared error (RMSE), mean squared error (MSE), and the error sum of squares (SSE). In (2), (3), and (4), are often used as the performance matrix, where \hat{y}_i is the predicted data, y_i is the current data, and n is the number of data samples [9]. In this research, the gradient descent was used as the learning rule. NARNN was chosen because LSTM and deep learning approaches require a complicated and careful design to produce accurate forecasts. In addition to that, these techniques work better with high dimensional and large datasets. Therefore, NARNN was selected in this research as the forecasting technique.

$$SSE = \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (2)$$

$$MSE = \frac{SSE}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_0 - Y_p)^2} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n |Y_0 - Y_p|}{n} \quad (5)$$

Y_0 = observed y_t

Y_p = Predicted y_t

The collected data was divided into training data and testing data. The training stage was used to test the model fit. Then, the time series forecasting model was established using the trained model. The performance was measured accordingly and then compared with actual values.

3.2. Local smoothing techniques

As discussed in section 1 the persistence of noise in a time series forecast can have continuously and cumulatively impair forecasting performance in n-steps ahead forecasts, so this issue has to be tackled carefully when working with forecasting algorithms while minimizing the effects of high or low frequency noise within the data, which can be useful for forecasting in the short- or long-term scale. The significance of noise

processing or removal was addressed in past work [7]-[11]. Next Section discusses various local smoothing techniques used in this paper.

3.2.1. Local regression techniques

The local regression method is based on the LOESS method [18]. It is based on fitting simple models to localized data subsets to form a curve that approximates the original data. The observations (x_i, y_i) are assigned neighborhood weights using the tricube weight function shown in (6). Let $\Delta_i(x) = |x_i - x|$ be the distance from x to x_i , and let $\Delta_i(x)$ be these distances in the smallest to largest order. Then, the neighborhood weight for the observation x_i, y_i is defined by the function $w_i(x)$:

$$w_i(x) = \left(\frac{\Delta_i(x)}{\Delta_q(x)} \right) \quad (6)$$

for x_i such that $\Delta_q(x) < \Delta_i(x)$, where q is the bandwidth that defines the number of observations in the subset of data localized around x . In the proposed algorithm, this approach was applied to fit a trend to the last k observations of resource utilization. Accordingly, a new trend line $\hat{g}(x) = \hat{a} + \hat{b}(x)$ is found for each new observation. This trend line is used to estimate the next observation $\hat{g}(x_k + 1)$. The new observation can be in the form of host resource utilization such as bandwidth slice utilization [18]. In (7) shows the final forecast formula using hybrid LOESS and NARNN:

$$y_t = \alpha_0 + \sum_{j=1}^k \alpha_j \varphi + \left(\sum_{i=1}^a \beta_{ij} (\hat{g}(x_k + 1))_{t-1} + \beta_{0j} \right) + \varepsilon_t \quad (7)$$

where a is the number of entries, k is the number of hidden layers with activation function φ , and β_{ij} is the parameter corresponding to the weight of the connection between the input unit i and the hidden unit j , α_j is the weight of the connection between the hidden unit and the output unit, and β_{0j} and α_0 are the constants that correspond, respectively, to the hidden unit j and the output unit. Two forms use LOWESS, which uses a first-degree polynomial model with weighted linear least squares and LOESS, which uses a second-degree polynomial model [18].

3.2.2. Robust local regression

This study adopted the LR method but the first fit was carried out with weights defined using the tricube weight function. The fit was evaluated at the x_i to get the fitted values (\hat{y}_i) , and the residuals $\hat{\varepsilon}_i = \hat{y}_i - y_i$, at each observation (x_i, y_i) , the additional robustness weight w_i was calculated, subjected to a magnitude of $\hat{\varepsilon}_i$. Accordingly, a new weight $w_i(x_i)$ was assigned to each observation, where w_i is defined as in (8) [18].

$$w_i = \begin{cases} \left(1 - \left(\frac{\hat{\varepsilon}_i}{6MAD} \right)^2 \right)^2, & |\hat{\varepsilon}_i| < 6MAD \\ 0, & |\hat{\varepsilon}_i| \geq 6MAD \end{cases} \quad (8)$$

where MAD is defined per (9):

$$MAD = Median(|\hat{\varepsilon}_i|) \quad (9)$$

Similarly, two versions were examined, i.e., 'RLOWESS' and 'RLOESS'. In both forms, the lower weights were assigned to the outliers in the regression. Moreover, outside the six mean absolute deviations, zero weights were assigned to new values.

3.2.3. Moving average

In several domains, time series data is usually smoothed using moving averages (MAs). This method is used especially in trend forecasting. The moving average is considered a type of real-time filter that removes high frequencies from data. In signal processing, MAs are therefore also called "low-pass filters" [19] where the calculated coefficients are equal to the reciprocal of the span or bandwidth. Moving averages are also known as "exponential smoothing". Let's define C_i as throughput at the time i . Let $c = \{C_i\}, i = 1 \dots p$ be the time series where p is the time series length. Therefore, the moving average of the period q at time l can be calculated as per (10) [19]. In (10) and (11) show the final forecast formula using hybrid moving average and NARNN:

$$m_l^q = \frac{1}{q} \sum_{i=1}^q c_{l-i+1} \quad (10)$$

$$y_t = \alpha_0 + \sum_{j=1}^k \alpha_j \varphi + (\sum_{i=1}^a \beta_{ij} ((m_l^q)_{t-1})_{t-1} + \beta_{0j}) + \varepsilon_t \quad (11)$$

3.2.4. Savitzky-Golay smoothing filter

The Savitzky-Golay (SG) smoothing filter is considered a type of low-pass filter characterized by two parameters denoted as K and M. The SG filter can be defined as a weighted moving average, i.e., a finite impulse response (FIR) filter. Filter coefficients are calculated using an un-weighted linear least-squares regression and a polynomial model of a specified degree (the default is 2). The time series to be estimated is denoted by $x(n)$, so the final output is obtained using (12), (13):

$$\hat{x}(n) = \sum_{k=-M}^M h(k)y(n-k) \quad (12)$$

$$y_t = \alpha_0 + \sum_{j=1}^k \alpha_j \varphi + (\sum_{i=1}^a \beta_{ij} ((\hat{x}(n))_{t-1})_{t-1} + \beta_{0j}) + \varepsilon_t \quad (13)$$

Note that a higher degree polynomial makes it possible to achieve a high level of smoothing without attenuating the data features [20]. It is worth mentioning that LOESS is used for seasonal decomposition, but in this work, the focus was to use LOESS and other local regression techniques as smoothening techniques, since decomposition may aggressively remove some of the important dataset features. Now, the question becomes how to select the bandwidth q . The bandwidth plays a critical role in the overall local regression fit; if the bandwidth selected is very small, large variances will result, as insufficient data will fall within the smoothing window, and, as a result, a noisy fit will be produced. On the other hand, if it is very large, not all data will be fitted within the specified window. Ideally, a separate bandwidth for each fitting point is used, bearing in mind features such as the local density. Practically, it is difficult to select an optimum q value, as the researcher does not want to unintentionally eliminate data. The simplest approach is to select q as a constant for all x_i . This case could be satisfactory for some simple constant variance data, but when the independent variables x_i have a non-uniform distribution such as in the bandwidth slice, problems such as empty neighborhoods and the accidental removal of more unnecessary data could result. Therefore, the following approach shown in Algorithm 1 was proposed:

Algorithm 1

```

Input y as time series bandwidth utilization
Output MSE
ŷ as a locally fitted (predicted) value using local smoothing techniques
1-Initialize, set q as 0
2-perform local smoothing using selected q
3-set q = q + 0.001
4-calculate the average MSE for all q-values
5-if MSE is = 0, then go to 2, else stop
6-set q ← q
7-return ŷ

```

Figure 1 shows the effects of different q -values and their corresponding differences from the original Bandwidth utilization. It is obvious that as the q -value increases, the smoother the curve, but the difference (error) will increase, in turn, increasing the overall absolute mean squared error (MSE). In this paper, NNAR (p, k) was used to indicate p lagged inputs and k nodes in the hidden layer. The general approach to searching for the optimal structure for the NNAR model is through trial-and-error, performed by testing numerous networks with varying numbers of inputs and hidden units and then calculating the generalization error of each to achieve a structure with the lowest generalization error [21], [22]. The crucial part of NNAR modeling is to find the appropriate values for p and k lagged inputs. In this work, Akaike's information criterion (AIC) [21]-[23] was used to automate the parameter selection process using R programming language. In fact, this method is asymptotically equivalent to cross-validation [23]. The best model with p and k was then chosen with the least value of AIC using the R language.

Two scenarios were examined in this paper-the short-term forecast, which shows how each hybrid technique will perform on the short-term scale, and the second scenario, which shows the forecast performance on a long-term scale forecast. Each time step represents 28.8 minutes and every 50 time steps represent one day. This case is due to the limitations in the data collection tool. The values were then interpolated, resulting in a time series model. The multi-scale forecast was used to investigate the extent to which the hybrid techniques would perform better than various forecast windows. The finding will prove beneficial for real-world core and backbone networks to achieve efficient network resource planning. In this paper, to enhance time series forecast models, the Box and Cox [24] power transformation was used to normalize series variances. Moreover, the augmented Dicky-Fuller (ADF) test [24] was used to confirm the stationarity of the time series although NARNN can be used to model a non-stationary time series. Previous work had advised examining the stationarity of regression models, as stationarity

could lead to misleading results [25]. To measure the stationarity of collected datasets, τ statistics was used. For a Type I test, the τ critical value is -2.985 when $n = 700$ (the datasets are composed of 700 data points). Since $\tau_{crit} = -2.223 < -10.14147 = \tau$, the null hypothesis that the time series is not stationary is rejected

4. RESULTS AND DISCUSSION

Figure 1 (a) shows the LTE bandwidth utilization without smoothing while Figures 1 (b) show the effects of applying moving average smoothing techniques using $q=0.002$ while Figure 1 (c) shows the effects of applying moving average smoothing techniques using $q=0.003$. As shown in Figure 1 (a), it is obvious that the bandwidth slice exhibited significant seasonal patterns with daily peaks. Nevertheless, the data also shows a stochastic pattern between successive points with continuous irregular fluctuations. On the other hand, no long-term trend appeared to exist. Minimum smoothing bandwidth (q) was selected as introduced in section 2 in algorithm 1. From Figure 1 (b), it is noticeable that the effects of applying smoothing techniques can be difficult to be observed by the naked eye. Therefore, MSE was accordingly calculated for each technique as depicted in Table 1, which shows the effect of applying various smoothing techniques on the selected dataset. Figure 1 (b) shows the LTE slice bandwidth utilization smoothed with the moving average (MA) and smoothing window $q=0.003$, which removes more of the small fluctuations at the top peaks, thus producing the highest MSE out of all the other techniques. In this case, (q) has a direct influence on the smoothing performance since it is inversely proportional to the MSE. Therefore, a significant portion of the data could be removed if higher (q) values were used. In fact, the higher the (q) values, the better the smoothing and the larger amount of data that will be lost as depicted in Figure 1 (c). Consequently, in today's data-centric world, losing even small amounts of data could lead to the violation of service level agreements in addition to inefficient resource utilization and planning. Therefore, (q) has to be selected according to algorithm (1). LOWESS produced the second largest MSE, as shown in Table 1, due to the likelihood that the nonlinear bandwidth slice would less likely fit if the first-degree polynomial linear model was used. However, fitting using the quadratic polynomial based on LOESS produced a smaller MSE, as shown in Table 1, due to the nonlinearity of the second-order local fitting models, as shown in Figure 1 (a). On the other hand, the sgolay filter produced a smaller MSE using a second-degree polynomial, in contrast to LOESS, which used a second-degree polynomial in which the weights were strongly influenced by the q bandwidth, as shown in (6). Finally, RLOESS and RLOWESS shared a similar performance, yielding the lowest MSE values, as shown in Table 1.

Table 1. The effects of applying algorithm 1

Smoothing Technique	q	Smoothing MSE
Moving average	0.003	2.4155e+07
LOWESS	0.005	2.0785e+07
LOESS	0.005	6.4096e+04
Sgolay	0.003	1.0133e-8
RLOWESS	0.002	1.7030e-10
RLOESS	0.002	1.7030e-10

Now, based on AIC calculated automatically from (autoarima) function in R, it was found that NARNN (28,14) produced the best fit. Table 2 shows the comparisons and the final results of applying the hybrid NARNN and smoothing techniques for the LTE bandwidth slice forecast for short 50 time steps head and for long 350 time steps ahead. Table 2 also shows the RMSE for NARNN of each smoothing technique for 50-time steps and 350-time steps. Overall, the hybrid NARNN tended to perform better, with better RMSE and a higher smoothing MSE.

It is worth to note that, the RMSE values when applying NARNN only without any combined technique were 308 for the 50-time step forecast and 323 for the 350-time step. From Table 2, it is obvious that the combination of LOESS and NARNN yielded better performance followed by the moving average and NARNN. The Diebold-Mariano test [21]-[23] was then applied to check for statistical significance. NARNN with LOESS RMSE was found to be statistically different from other hybrid techniques. The same finding was found for the 350 time step forecast. Therefore, NARNN with LOESS yielded better performance and was verified statistically via the Diebold-Mariano test as well. This result confirms the effectiveness and the reliability of the hybrid NARNN and the smoothing techniques for forecasting short- and long-term scales. The autocorrelation function (ACF) obtained using the Ljung-Box test was used for further analysis. The analysis of ACF was used to calculate the number of inputs of auto-correlated vectors to create an appropriate model. Moreover, it was also used to investigate white noise (zero mean, constant variance, uncorrelated processes, and normally distributed) in the residuals. Figure 2 (a) shows the ACF and the plots of the residuals of the hybrid NARNN smoothing forecast models for the 50-time step. And Figure 2 (b) for 350-time step ahead.

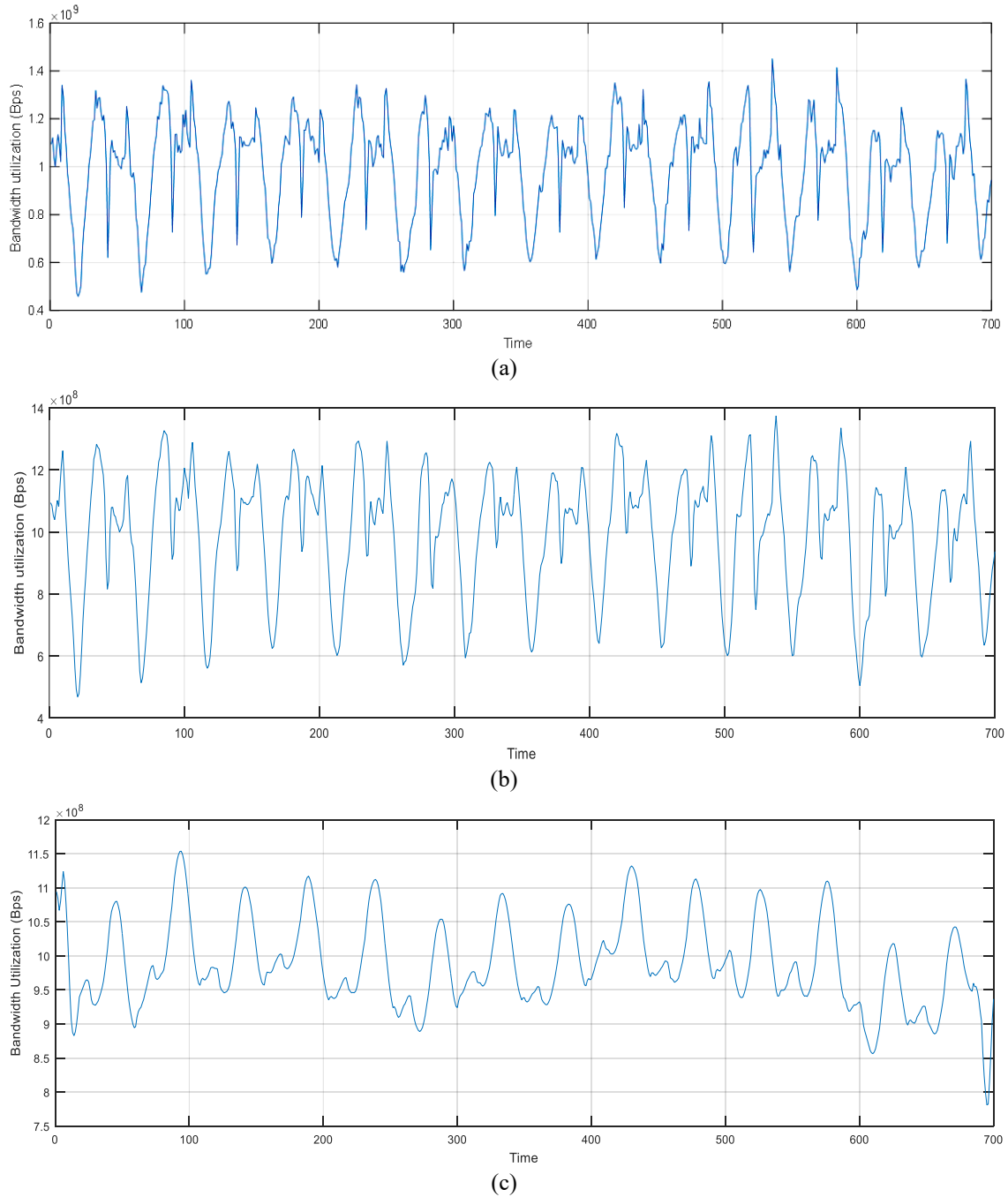


Figure 1. Bandwidth utilization using moving average: (a) original LTE slice, (b) LTE slice smoothed with $q=0.003$, and (c) LTE slice smoothed with $q=0.05$

Table 2. RMSE results of applying the hybrid techniques

Smoothing Technique	q	RMSE of 50-time steps ahead forecast using NARNN + smoothing	RMSE of 350-time steps ahead forecast using NARNN + smoothing
Moving Average	0.003	272	295
LOWESS	0.005	289	330
LOESS	0.005	270	293
Sgolay	0.003	309	351
RLOWESS	0.002	292	340
RLOESS	0.002	289	298

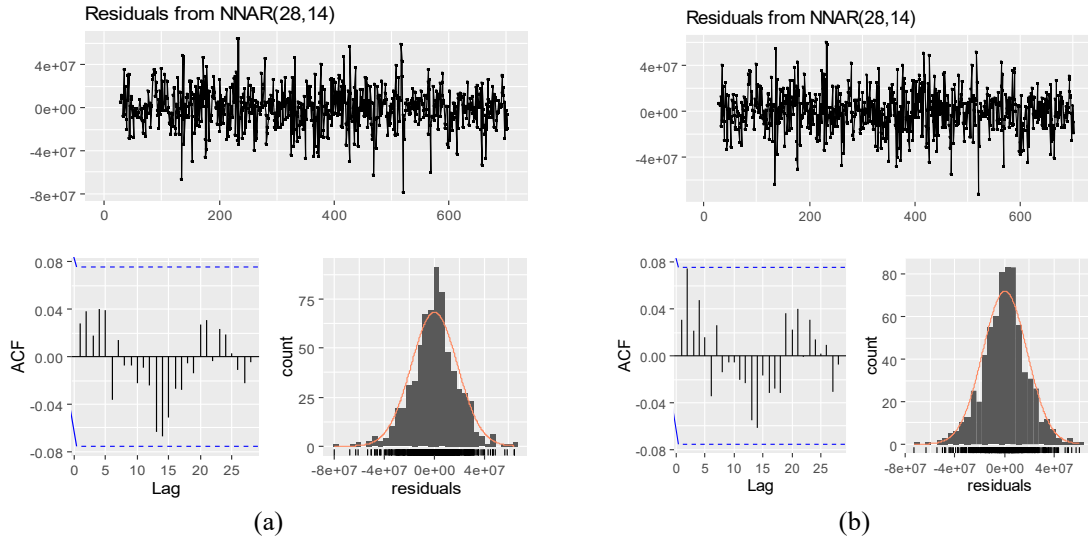


Figure 2. The ACF and the plots of the residuals of the hybrid NARNN smoothing forecast models for: (a) the 50-time step and (b) 350-time step

In the case of the 50-time step forecast using the hybrid NARNN with LOESS, the residuals fell randomly within the horizontal band (between $4e7$ and $-4e7$) and as a result the variance of the residuals looked to be independent of the size of the fitted values. Meanwhile, the same results were found for 350-time steps forecast in hybrid NARNN. This pattern suggests that the variances in the error terms are equal. Moreover, no one residual stood out from the random pattern; thus, suggesting that there were no outliers. The lags in the ACF plots fell below the 0.08 threshold. Moreover, no pattern was evident in the residuals. Additionally, the residuals followed a random distribution around zero. This result confirms and validates that the NARNN with LOESS relatively provided the best forecasting models. Figure 3 (a) shows the performance comparison of hybrid NARNN versus hybrid Seasonal Autoregressive Moving Average (SARIMA) for 50-time step forecast, the hybrid SARIMA was used as a benchmark to validate the obtained results. Results have shown that in overall NARNN hybrid technique outperform other non-hybrid techniques. NARNN-LOESS had the least RMSE values across other SARIMA hybrid techniques, although SARIMA-original slightly outperform NARNN when used without local smoothing techniques. Therefore, NARNN-LOESS will be our best choice since our objective is to provide best forecast performance with minimum data lose as discussed earlier. Same findings were found for the 350-time step forecast as depicted in Figure 3 (b).

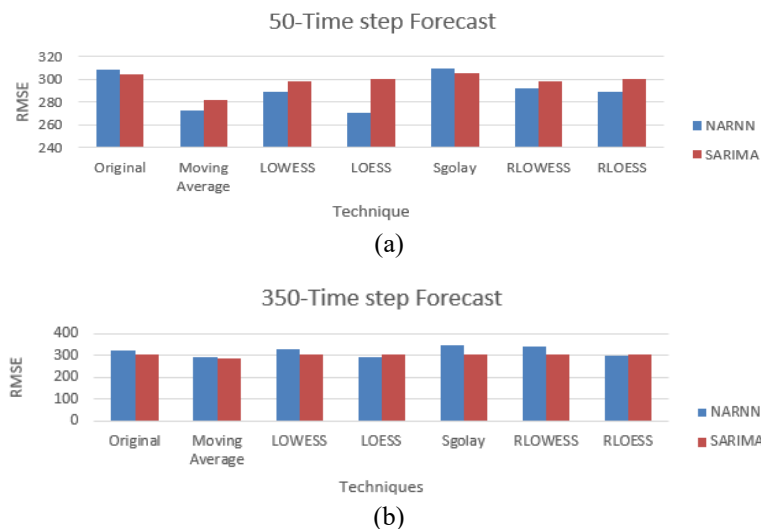


Figure 3. Performance comparison of hybrid NARNN versus SARIMA: (a) for 50-time step forecast and (b) for 350-time step forecast

In this case, NARNN-LOESS had the least RMSE value compared to SARIMA with hybrid techniques. Although SARIMA hybrid show a noticeable performance improvement compared to hybrid NARNN except for the case of hybrid NARNN-LOESS that barely outperform SARIMA-LOESS. The Diebold-Mariano test was then applied to check the statistical significance of the obtained results. It was found that RMSE of NARNN-LOESS hybrid techniques to be better and statistically different from forecasting SARIMA and this confirms the superiority of the NARNN hybrid techniques. Figure 4. (a) depicts the 50-time step ahead forecast for the NARNN with LOESS and Figure 4 (b) for 350-time step forecast, both figures show that the both forecast can effectively lie between the prediction intervals.

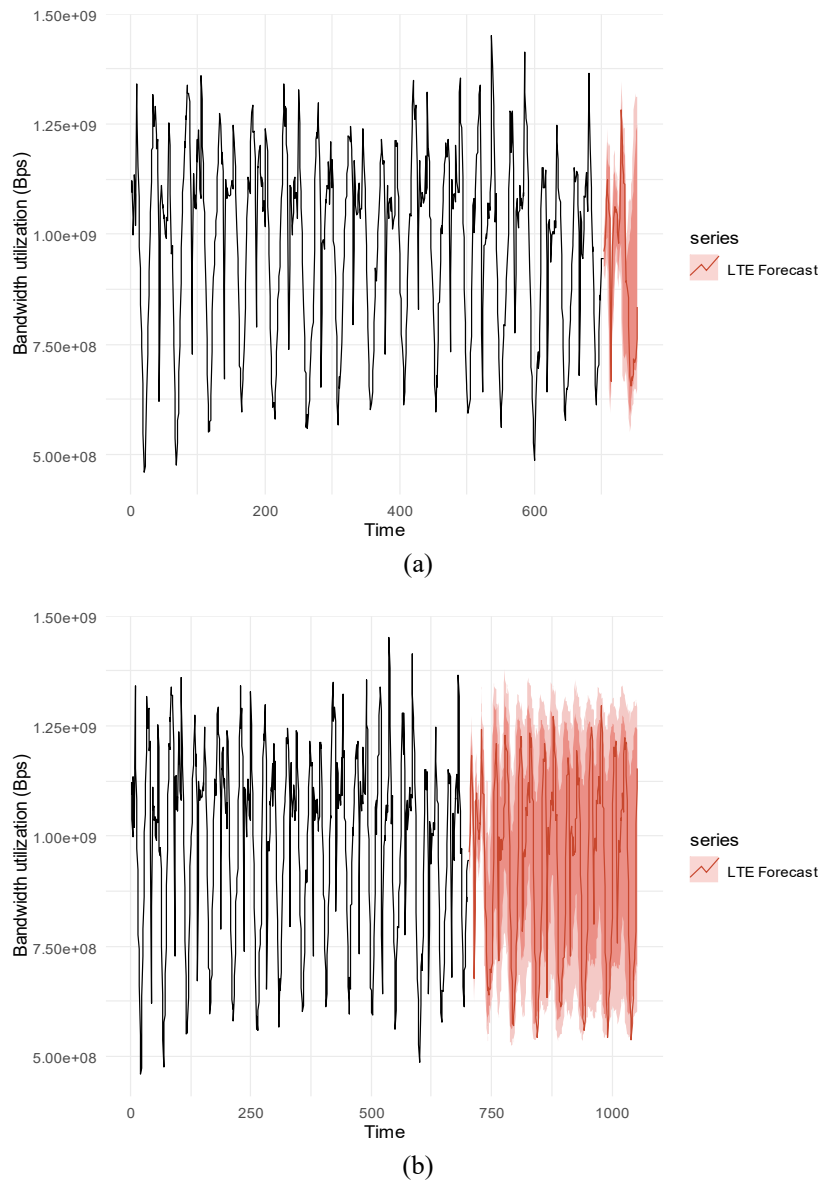


Figure 4. (a) Time series forecast for the NARNN with LOESS: (a) For 50 Time steps and (b) for 350-time

5. CONCLUSION

In this paper, hybrid local smoothing and neural network auto-regressive (NNAR) modeling approaches were used to forecast LTE core bandwidth slice utilization. Several local smoothing techniques were analyzed, and a local smoothing mechanism was introduced to minimize the effects of data losses, which may carry necessary information resulting from aggressive and uncontrolled smoothing functions. The models showed better forecast performance in terms of RMSE, provided minimum data losses were maintained. Long-term and short-term step forecasts were examined and the results were verified using residual analysis,

performance comparison and statistical significance tests. The proposed method can be used for slice traffic forecast in 5G slice resource forecast and management, as well current and future backbone networks.

REFERENCES

- [1] R. Boutaba, M. Salahuddin, N. Limam, and S. Ayubi, "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *J Internet Serv Appl*, vol. 9, no. 16, 2018, doi: 10.1186/s13174-018-0087-2.
- [2] M. Oshi and T. H. Hadi, "A review of network traffic analysis and prediction techniques," *arXiv preprint*, 2015.
- [3] P. Cortez, M. Rio, M. Rocha and P. Sousa, "Internet Traffic Forecasting using Neural Networks," *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, Vancouver, BC, Canada, 2006, pp. 2635-2642, doi: 10.1109/IJCNN.2006.247142.
- [4] S. Chabaa, A. Zeroual, and J. Antari, "Identification and Prediction of Internet Traffic Using Artificial Neural Networks," *Journal of Intelligent Learning Systems and Applications*, vol. 02, no. 03, pp. 147-155, 2010, doi:10.4236/jilsa.2010.23018.
- [5] Y. Zhu, G. Zhang, and J. Qiu, "Network Traffic Prediction based on Particle Swarm BP Neural Network," *J. Networks*, vol. 8, no. 11, pp. 2685-2691, 2013.
- [6] Y. Li, H. Liu, W. Yang, D. Hu and W. Xu, "Inter-data-center network traffic prediction with elephant flows," *NOMS 2016 - 2016 IEEE/IFIP Network Operations and Management Symposium*, Istanbul, Turkey, 2016, pp. 206-213, doi: 10.1109/NOMS.2016.7502814.
- [7] S. Dyllon, T. Hong, O. Oumar, P. Xiao, "Educational bandwidth traffic prediction using non-linear autoregressive neural networks," *Conference: CLAWAR 2018*, Sep. 2018.
- [8] W. Yoo, and A. Sim, "Time-Series Forecast Modeling on High-Bandwidth Network Measurements," *J Grid Computing*, vol. 14, no. 3, pp. 463-476, 2016, doi: 10.1007/s10723-016-9368-9.
- [9] D. Afolabi, S. U. Guan, K. L. Man, P. W. H. Wong, "Hierarchical Meta-Learning in Time Series Forecasting for Improved Interference-Less Machine Learning," *Symmetry*, vol. 9, no. 11, 2017, doi: 10.3390/sym9110283.
- [10] T. W. Joo, and S. B. Kim, "Time Series Forecasting Based on Wavelet Filtering," *Expert Systems with Applications*, vol. 42, no. 8, pp. 3868-74, 2015, doi: 10.1016/j.eswa.2015.01.026.
- [11] B. Doucoure, K. Agbossou, and A. Cardenas, "Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data," *Renewable Energy*, vol. 92, pp. 202-211, 2016, doi: 10.1016/j.renene.2016.02.003.
- [12] I. Alawe, A. Ksentini, Y. Hadjadj-Aoul and P. Bertin, "Improving Traffic Forecasting for 5G Core Network Scalability: A Machine Learning Approach," *IEEE Network*, vol. 32, no. 6, pp. 42-49, 2018, doi: 10.1109/MNET.2018.1800104.
- [13] J. Wang, Z. Wang, J. Li, and J. Wu, "Multilevel Wavelet Decomposition Network for Interpretable Time Series Analysis," *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, doi: 10.1145/3219819.3220060.
- [14] S. M. Salih, "Bandwidth Utilization Prediction in LAN Network Using Time Series Modeling," *Iraqi Journal of Computer, Communication, Control and System Engineering*, pp. 78-89, 2019, doi: 10.33103/uo.ijccce.19.2.9.
- [15] J. Feng, X. Chen, R. Gao, M. Zeng and Y. Li, "DeepTP: An End-to-End Neural Network for Mobile Cellular Traffic Prediction," *IEEE Network*, vol. 32, no. 6, pp. 108-115, 2018, doi: 10.1109/MNET.2018.1800127.
- [16] L. Le, D. Sinh, L. Tung and B. P. Lin, "A practical model for traffic forecasting based on big data, machine-learning, and network KPIs," *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, Las Vegas, NV, USA, 2018, pp. 1-4, doi: 10.1109/CCNC.2018.8319255.
- [17] J. You, H. Xue, L. Gao, G. Zhang, Y. Zhuo and J. Wang, "Predicting the online performance of video service providers on the internet," *Multimedia Tools and Applications*, vol. 76, no. 18, pp. 19017-19038, 2017, doi: 10.1007/s11042-017-4460-0.
- [18] W. S. Cleveland, "LOWESS: A Program for Smoothing Scatterplots by Robust Locally Weighted Regression," *The American Statistician*, vol. 35, no. 1, pp. 829-836, 1981, doi:10.2307/2683591.
- [19] A. Raudys and Ž. Pabarškaitė, "Optimising the Smoothness and Accuracy of Moving Average for Stock Price Data," *Technological and Economic Development of Economy*, vol. 24, no. 3, pp. 984-1003, 2018, doi: 10.3846/20294913.2016.1216906.
- [20] J. Luo, K. Ying, and J. Bai, "Savitzky-Golay smoothing and differentiation filter for even number data," *Signal Processing*, vol. 85, no. 7, pp. 1429-1434, 2005, doi: 10.1016/j.sigpro.2005.02.002.
- [21] A. Maleki, S. Nasser, and M. S. Aminabad, "Comparison of ARIMA and NNAR Models for Forecasting Water Treatment Plant's Influent Characteristics," *8KSCE Journal of Civil Engineering*, vol. 22, no. 6, pp. 1-13, 2018, doi: doi.org/10.1007/s12205-018-1195-z.
- [22] S. Bhattacharya, A. Biswas, S. Nandi, and S. K. Patra, "Exhaustive model selection in $b \rightarrow sll$ decays: Pitting cross-validation against the Akaike information criterion," *Physical Review D*, vol. 101, no. 5, 2020.
- [23] H. Abdollahi, "A novel hybrid model for forecasting crude oil price based on time series decomposition," *Applied Energy*, vol. 267, no. 8, 2020, doi: 10.1016/j.apenergy.2020.115035
- [24] T. Kulaksizoglu, "Lag Order and Critical Values of the Augmented Dickey-Fuller Test: A Replication," *Journal of Applied Econometrics*, vol. 30, no. 6, pp. 1010-1010, 2015.

- [25] S. Bunrit, N. Kerdprasop and K. Kerdprasop, "Multiresolution Analysis Based on Wavelet Transform for Commodity Prices Time Series Forecasting," *International Journal of Machine Learning and Computing*, vol. 8, no. 2, pp. 175-180, 2018, doi: 10.18178/ijmlc.2018.8.2.683.

BIOGRAPHIES OF AUTHORS



Mohamed khalafalla Hassan is a researcher and ICT specialist with 16 years of wide range of research and ICT experience. Received his BSc in computer engineering in 2004 from future university Sudan and MSc from University Putra Malaysia (UPM) in 2009 in communication network engineering, currently he is a PhD candidate in communication engineering at university technology Malaysia (UTM), his interests are but not limited to software defined networks, application of Machine learning in communication networks and Network Function Virtualization.



Sharifah Hafizah Sayed Ariffin Received her B.Eng. (Hons) from London, in 1997, and obtained her MEE and Ph.D. in 2001 and 2006 from Universiti Teknologi Malaysia, and Queen Mary, University of London, London respectively. She is currently an Associate Professor with the Faculty of Electrical Engineering, Universiti Teknologi Malaysia. Her current research interest are on Internet of Things, Ubiquitous computing and smart devices, Wireless sensor networks, IPV6, Handoff Management, Network and Mobile Computing System. She had published 116 papers, 17 copyrights, 1 integrated circuit, and 1 trademark.



Sharifah Kamilah Syed-Yusof received B.Sc. degree in electrical engineering from George Washington University, USA in 1988 and obtained her M.E.E. and Ph.D. degrees from UTM in 1994 and 2006, respectively. She is currently working as a full professor with Faculty of Electrical Engineering, UTM. Her research interest includes wireless communication, software defined radio and cognitive radio. She is a member of Eta Kappa Nu (HKN).



Nurzal Effiyana Binti Ghazali received MS. degree in electrical engineering from Shibaura Institute of Technolog and Ph.D. degrees from UTM in 2016, she is doing research in mobile computing, mobility management, network communication protocol, and sport monitoring system.



Mohammed EA Kanona received his BSc, MSc and PhD from Future University Sudan, currently he is assistant professor at Future university, his current interests are, application of signal processing and machine learning in communication networks.