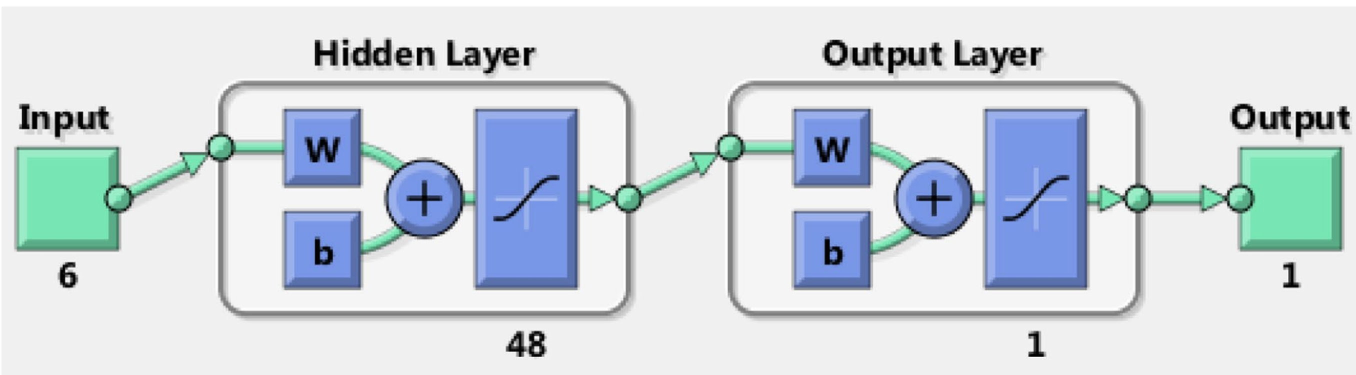
cogent engineering



ELECTRICAL & ELECTRONIC ENGINEERING | RESEARCH ARTICLE

Optimal model for path loss predictions using feed-forward neural networks

Segun I. Popoola, Emmanuel Adetiba, Aderemi A. Atayero, Nasir Faruk and Carlos T. Calafate

Cogent Engineering (2018), 5: 1444345

cogent engineering

ELECTRICAL & ELECTRONIC ENGINEERING | RESEARCH ARTICLE

Optimal model for path loss predictions using feed-forward neural networks

Segun I. Popoola[1*], Emmanuel Adetiba[1,2], Aderemi A. Atayero[1], Nasir Faruk[3] and Carlos T. Calafate[4]

Received: 16 November 2017
Accepted: 20 February 2018
First Published: 26 February 2018

*Corresponding author: Segun I. Popoola, Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria
E-mails: segun.popoola@covenantuniversity.edu.ng, segunpopoola15@gmail.com

Reviewing editor:
Chau Yuen, Singapore University of Technology and Design, Singapore

Additional information is available at the end of the article

Abstract: In this paper, an optimal model is developed for path loss predictions using the Feed-Forward Neural Network (FFNN) algorithm. Drive test measurements were carried out in Canaanland Ota, Nigeria and Ilorin, Nigeria to obtain path loss data at varying distances from 11 different 1,800 MHz base station transmitters. Single-layered FFNNs were trained with normalized terrain profile data (longitude, latitude, elevation, altitude, clutter height) and normalized distances to produce the corresponding path loss values based on the Levenberg–Marquardt algorithm. The number of neurons in the hidden layer was varied (1–50) to determine the Artificial Neural Network (ANN) model with the best prediction accuracy. The performance of the ANN models was evaluated based on different metrics: Mean Absolute error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), standard deviation, and regression coefficient (R). Results of the machine learning processes show that the FNN architecture adopting a tangent activation function and 48 hidden neurons produced the least prediction error, with MAE, MSE, RMSE, standard deviation, and R values of 4.21 dB, 30.99 dB, 5.56 dB, 5.56 dB, and 0.89, respectively. Regarding generalization ability, the predictions of the optimal ANN model yielded MAE, MSE, RMSE, standard deviation, and R values of 4.74 dB, 39.38 dB, 6.27 dB, 6.27 dB, and 0.86, respectively, when tested with new data not previously included

Segun I. Popoola

ABOUT THE AUTHOR

Segun I. Popoola is currently at the verge of completing his Master of Engineering (MEng) program in Information and Communication Engineering at the Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria. He graduated from Ladoke Akintola University of Technology, Ogbomoso, Nigeria with a Bachelor of Technology (*First Class*) degree in Electronic and Electrical Engineering. Segun was awarded the best graduated student in the Department of Electronic and Electrical Engineering by the Faculty of Engineering and Technology (FET) in conjunction with the Nigerian Society of Engineers (NSE). He has authored and co-authored more than 30 academic papers published in international peer reviewed journals and conference proceedings. He is a member of International Association of Engineers (IAENG). His research interests are in, but not limited to: wireless communications, machine learning, radio propagation modeling, Internet of Things (IoT); cloud computing, and Smart and Connected Cities.

PUBLIC INTEREST STATEMENT

The huge potentials of Information and Communication Technology can be leveraged for the timely attainment of the Sustainable Development Goals. To this end, global interconnectedness is viewed as an enabling technological platform for digital transformation. Unfortunately, the coverage of mobile connectivity that is expected to drive digital inclusion is not yet global. Path loss prediction models are vital tools for radio coverage estimation, determination of base station location, frequency allocation, antenna selection, and interference feasibility studies during radio network planning. Machine learning techniques may be exploited for path loss predictions in rural and urban propagation environments. In this paper, an optimal model is developed for path loss predictions using Feed-Forward Neural Network algorithm. The performance of the developed ANN model proved to be optimal in terms of prediction accuracy and generalization ability when compared to those of widely used empirical models (Hata, COST 231, ECC-33, and Egli).

in the training process. Compared to the Hata, COST 231, ECC-33, and Egli models, the developed ANN model performed better in terms of prediction accuracy and generalization ability.

Subjects: Neural Networks; Communication Technology; Electromagnetics & Communication

Keywords: path loss; received signal strength; scale conjugate gradient; radio network planning; Artificial Neural Network

1. Introduction

The huge potentials of Information and Communication Technology (ICT) can be leveraged for the timely attainment of the Sustainable Development Goals (Armenta, Serrano, Cabrera, & Conte, 2012). To this end, global interconnectedness is viewed as an enabling technological platform for digital transformation. Unfortunately, the coverage of mobile connectivity that is expected to drive digital inclusion is not yet global. In fact, sub-Saharan Africa remains the most under-penetrated region of the world (GSMA, 2014). Extending mobile network coverage to those who are furthest behind will help in bridging the wide digital divide between rural and urban areas, and provide the enabling technology and infrastructure for ICT-driven applications and services (Popoola, Atayero, Okanlawon, Omopariola, & Takpor, 2018). Consequently, public services such as health care and education will become more accessible and affordable (Matthews, Osuoyah, Popoola, Adetiba, & Atayero, 2017; Popoola, Atayero, Badejo, et al., 2018). Harnessing this golden opportunity, economies of developing countries can leapfrog in the areas of agriculture, e-commerce, and transportation.

Meanwhile, the need for greater cellular network capacity will exponentially increase the rate of deployment of base stations, making the determination of suitable locations more difficult. More so, the design of mobile communication networks requires a good knowledge of the wireless channel (Faruk, Adediran, & Ayeni, 2013; Oseni, Popoola, Enumah, & Gordian, 2014). It largely determines the transmission rate and the quality of signal propagation due to its complexity and randomness (Sotiroudis & Siakavara, 2015). Interactions between radiated electromagnetic waves and physical objects in wireless propagation environment often result in reflection on large plane surfaces, scattering from surfaces of small size relative to the wavelength of transmission, transmission through dense materials like walls or floors, or shadowing by obstacles such as buildings and foliage. Therefore, radio waves that are transmitted by the base station antennas reach mobile devices by different propagation mechanisms, depending on the environment. This often results in signal fading, which may be in small scale or large scale (Rappaport, 1996). Small-scale signal fading occurs due to rapid fluctuations of received signal strength over a short period of time and small distance (Phillips, Sicker, & Grunwald, 2013); instead, large-scale signal fading takes place as average signal strength changes over a large distance between the base station and the mobile station (Faruk, Adediran, & Ayeni, 2013). The effect of large-scale fading is also known as *path loss*.

Path loss prediction models are vital tools for radio coverage estimation, determination of base station location, frequency allocation, antenna selection, and interference feasibility studies during radio network planning (Popoola, Badejo, Ojewande, & Atayero, 2017). Prior to actual mobile network deployment, radio engineers use these models to understand wireless channel characteristics and to predict signal attenuation. Propagation models can be broadly organized into two categories, namely: deterministic and empirical models. Deterministic models are based on theoretical principles of diffraction (Luebbers, 1984), ray tracing (Mohtashami & Shishegar, 2012), integral equation (Hufford, 1952), and parabolic equation (Zelley & Constantinou, 1999); empirical models are based on practical measurements conducted in a particular environment. Although deterministic models are more accurate, they lack computational efficiency. Empirical models such as Okumura-Hata model (Hata, 1980), COST 231 model (Erceg, 1999), and standard propagation model (Popoola & Oseni, 2014b) are easy to implement with satisfactory computational efficiency in terms of time and cost. However, they are not as accurate as deterministic models because they do not effectively account for the unique geographical configurations of the propagation environment. Meanwhile, the reliability of the radio access network depends on the accuracy of the propagation model employed. Hence, the need for significant improvement in the prediction accuracy of empirical models while maintaining model simplicity and ease of use.

Radio propagation environments have been widely categorized into rural, suburban, and urban (Rappaport, 1996). These environments are composed of varying unique geographical features with different altitude, terrain height information, land usage data, building shape and height information, and building surface characteristics. Of all available empirical models, previous research works (Al Salameh & Al-Zu'bi, 2015; Faruk, Ayeni, & Adediran, 2013; Ibhaze, Ajose, Atayero, & Idachaba, 2016; Nimavat & Kulkarni, 2012; Oseni, Popoola, Abolade, & Adegbola, 2014; Popoola & Oseni, 2014a; Rath, Verma, Simha, & Karandikar, 2016) have identified the Okumura-Hata, COST 231-Hata, and SPM path loss prediction models as appropriate for radio network planning in the 1,800 MHz band. These models accounts for propagation path loss based on radio parameters including heights of transmitter and receiver antennas, frequency of transmission, and distance between the base station and the mobile station. However, the presence of various sources of clutter in the propagation environment contributes largely to propagation path loss (Oseni, Popoola, Enumah, et al., 2014). Modeling large scale channel fading without consideration for altitude, land use, and clutter height results in path loss predictions with large deviation from real measurement values.

Okumura-Hata model is an empirical formulation of the graphical path loss data that was collected at 150–1,500 MHz band (Hata, 1980). The separation distance between the transmitter and the receiver ranges from 1 to 20 km. The appropriateness of the empirical model for path loss prediction in practical environments has been widely investigated in the literature. Medeisis and Kajackas (2000) investigated the suitability of Okumura-Hata model for path loss prediction in different Very High Frequency (VHF) and Ultra-High Frequency (UHF) bands. Although the empirical model performed fairly well in a built-up environment, the prediction error was significant in a rural propagation environment. A least square technique was applied to the model to reduce the high prediction error. The prediction accuracy of Okumura-Hata model was enhanced in Schneider, Lambrecht, and Baier (1996) with the details of the morphology and buildings in the wireless channel. The findings of the authors showed that better prediction accuracy will be obtained if morphological data and building data are incorporated into the model for rural/suburban and urban environments respectively. Farhoud, El-Keyi, and Sultan (2013) examined the applicability of Okumura-Hata model in the Global System for Mobile Communications (GSM) 900 MHz band and introduced correction factors to improve the accuracy of the model for different regions in Egypt. Akhoondzadeh-Asl and Noori (2007) suggested another way of defining the antenna height of the base station in Okumura-Hata model. The empirical model was adapted for path loss predictions by performing a cubic regression on field measurement data that was collected in Salalah, Oman (Nadir & Ahmad, 2010). Mardeni and Pey (2010) optimized Okumura-Hata model for urban outdoor coverage in the Code Division Multiple Access (CDMA) system in Malaysia. Major differences were found in the parameters of Okumura-Hata model when it was applied to railway environment at 900 MHz and this findings were reported in Cota, Serrador, Vieira, Beire, and Rodrigues (2013). Begovic, Behlilovic, and Avdic, (2012) evaluated the applicability of a set of empirical models for WiMAX coverage planning at 3.5 GHz. Adeyemo, Ogunremi, and Ojedokun (2016) optimized Okumura-Hata model for LTE signal attenuation in Lagos, Nigeria using the Least Square method. The model was also tuned for TErrestrial Trunked RAdio (TETRA) mobile radio applications in Saudi Arabia as reported in Alamoud and Schütz (2012).

COST 231 extends Okumura-Hata model to cover the frequency range of 1,500–2,000 MHz (Erceg, 1999). The transmitter antenna height and the receiver antenna height can be in the range of 30–200 m and 1–10 m respectively. In the study conducted in dense urban areas at 1,800 MHz by Verma and Saini (2016), COST 231 model had the lowest RMSE with the most acceptable standard deviation when compared to free space, SUI, and ECC models. In addition, SPM was developed based on the Hata path loss formulas (Popoola, Atayero, Faruk, Calafate, Adetiba, et al., 2017; Popoola, Atayero, Faruk, Calafate, Olawoyin, et al., 2017). It determines the large-scale fading of received signal strength over a distance range of 1–20 km. Therefore, it is appropriate for mobile channel characterization of popular cellular technologies such as GSM. Although distance is usually expressed in *km* in Hata formulas, SPM accepts distance values in meters. SPM ignored the effects of diffraction, clutter, and terrain. It assumed that appropriate settings of the parameters which account for only one clutter class will cater for the influence of these external factors on signal propagation. The correction function for the mobile receiver antenna height was also ignored for $h_r \leq 1.5$ m since it has negligible values for an average mobile antenna height.

Machine learning techniques may be exploited for path loss predictions in rural and urban propagation environments (Salman et al., 2017). ANN is an adaptive statistical tool that models the biological nervous system to solve regression problems. The capability of ANNs to model complex nonlinear functional relationships provides an opportunity to improve the accuracy of empirical path loss models with better computational efficiency. In this paper, an optimal model is developed for path loss predictions using a Feed-Forward Neural Network (FFNN) algorithm. Drive test measurements are carried out in Canaanland Ota, Nigeria (Latitude 6°40′30.3″N, Longitude 3°09′46.3″E) and Ilorin, Nigeria (Latitude 8°29′12.5″N, Longitude 4°30′23.3″E) to obtain path loss data at varying distances from 11 different 1,800 MHz base station transmitters. Single-layered FFNNs are trained with normalized terrain profile data (longitude, latitude, elevation, altitude, clutter height) and normalized distances to produce corresponding path loss values based on the Levenberg–Marquardt algorithm. The number of neurons in the hidden layer is varied (1–50) to determine the Artificial Neural Network (ANN) model with the best prediction accuracy. The performance achieved by the different ANN models is evaluated based on metrics such as Mean Absolute error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), standard deviation, and regression coefficient (*R*).

The rest of the paper is organized as follows: Section 2 explains the data collection process and the development stages of the ANN model; Section 3 discusses the results of the experimentations; and finally, Section 4 concludes the paper and discusses future works.

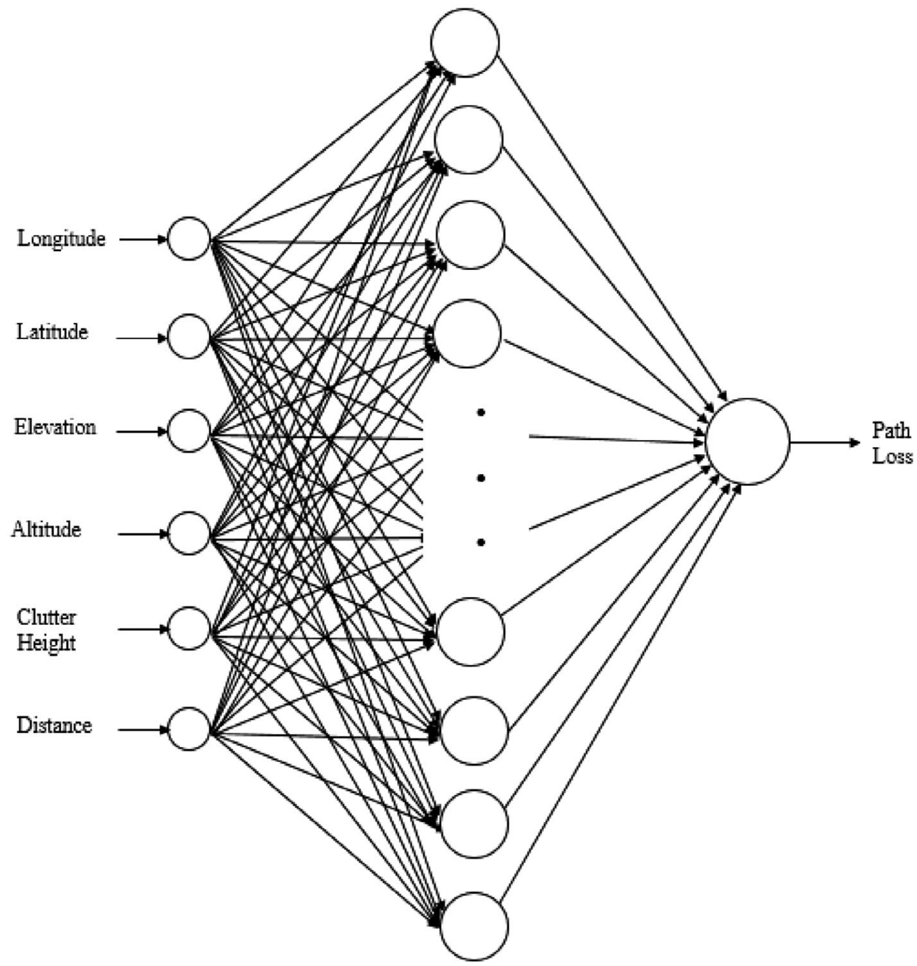
2. Materials and method

2.1. Field measurement and data collection

Extensive measurement campaigns were conducted in Canaanland Ota, Ogun State, Nigeria (Latitude 6°40′30.3″N, Longitude 3°09′46.3″E) and Ilorin, Kwara State, Nigeria (Latitude 8°29′12.5″N, Longitude 4°30′23.3″E) to obtain path loss data at varying distances from eleven different 1,800 MHz base station transmitters. Two different locations were selected to accommodate sufficient diversity in the propagation environment. Typical urban, suburban, and rural propagation environments were covered by the planned survey routes. A total of eleven survey routes (R1–R11) were mapped out to cover radio wave propagation in the antenna direction of each of the base station transmitters. Survey routes R1–R8 are located within Canaanland Ota, Ogun State, Nigeria, while the remaining three (R9–R11) are within Ilorin, Kwara State, Nigeria. Typical urban, suburban, and rural propagation environments were covered by the planned survey routes.

Radio Frequency (RF) signal measurements were carried out by drive test under good climatic conditions. Also, good vehicular accessibility to base station locations were considered for a smooth test drive. Distances covered by the drive routes are considered long enough to allow the noise floor of the receiver to be reached. The data collection process was performed with the use of the Transmission Evaluation and Monitoring System (TEMS) network performance investigation software (Popoola, Atayero, & Faruk, 2018; Popoola, Atayero, Faruk, & Badejo, 2018). TEMS Investigation has data collection, real-time network data analysis, and post-data processing capabilities. This network testing software ran on an Intel Core i5–3210MCPU@2.50 GHz speed with 4 GB Random Access Memory (RAM) and 64-bit Windows 7 operating system. A TEMS mobile station, the software USB dongle, and a Garmin Global Positioning System (GPS) were connected to the laptop. The whole setup was carefully placed in a vehicle, and the vehicle was driven at an average speed of 40 km/h. This

Figure 1. ANN model architecture.



speed was maintained to minimize Doppler effects. The data collected were pre-processed in Microsoft Excel 2016.

2.2. ANN model design and development

A single-layered feed-forward neural network architecture, comprised of six input neurons and one output neuron, was designed for model training and development, as shown in Figure 1. The pre-processed path loss data obtained through drive test measurements along the eleven survey routes were combined, sorted, and later divided into training data-set and testing data-set. The training and testing datasets are stored in matrix form with 9,074 and 1,297 instances, respectively. Testing data-set comprised of data instances that are not included in the training data-set. The testing data-set was created to evaluate the generalization ability of the ANN models. The input data variables of the datasets include longitude, latitude, elevation, altitude, clutter height, and distance. The single target output of the ANN model is the corresponding path loss value for the specified input vector matrix.

A minimum-maximum normalization process was performed on the input matrices of both datasets to prevent impulsive changes due to large variation in the datasets (Popoola, Misra, & Atayero, 2018). The scaling of the data prevented larger values from overshadowing the smaller values. It also aided the learning process by avoiding untimely saturation of hidden nodes. The min-max normalization equation is given by Equation (1).

$$p_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}}(m - n) + n \tag{1}$$

where

p_i = data range of four input values

m = maximum normalized value = 0.9

n = minimum normalized value = −0.9

$y_{\min}$ = minimum values of the activated function

$y_{\max}$ = maximum values of the activated function

In order to determine the ANN model with the best prediction accuracy, the number of neurons in the hidden layer was varied from 1 to 50. The ANN models were trained based on the Levenberg–Marquardt learning algorithm. A non-linear relationship was established between the normalized input variables and the single output variable based on learning rule. Neural network model design, training, validation, and testing were all done using Machine Learning and Statistics Toolbox available in MATLAB 2017a produced by MathWorks Inc. The learning approach used was the supervised back propagation owing to its popularity and ease of learning. The path losses were determined by the values of the input variables, and the errors generated in the process of supervised learning were approximated. For the artificial neurons to approximate functions in a flexible manner, a linear activation function was employed at the input layer. At the hidden layer, the input data were transformed into non-linear form using the hyperbolic tangent sigmoid activation function as represented by Equation (2).

$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \tag{2}$$

The most suitable number of neurons in the hidden layer was determined through experimentation. The complete training data-set was divided into 70% training, 15% validation, and 15% testing sub-datasets in the ANN model design (Adetiba, Iweanya, Popoola, Adetiba, & Menon, 2017; Adetiba & Olugbara, 2015). For each instance of number of hidden neurons, the model training was performed ten times to account for any inconsistency. The prediction accuracies and the generalization ability of the ANN models were evaluated based on different metrics such as MAE, MSE, RMSE, standard deviation, and R, relative to the target values in the training and testing datasets, respectively. The mathematical expressions of these statistical performance metrics are given by Equations (3)–(7).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}(\text{PL}_i^m - \text{PL}_i^p) \tag{3}$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\text{PL}_i^m - \text{PL}_i^p)^2 \tag{4}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\text{PL}_i^m - \text{PL}_i^p\right)^2} \tag{5}$$

$$\text{SD} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(|\text{PL}_i^m - \text{PL}_i^p| - \mu\right)^2} \tag{6}$$

Table 1. Mean terrain profile data of survey routes

Route	Longitude	Latitude	Elevation (m)	Altitude (m)	Clutter height (m)	Distance (m)	Path loss (dB)
R1	3.1651	6.6777	54.34	59.81	4.95	409.47	142.81
R2	3.1672	6.6742	54.18	61.39	4.92	496.65	140.30
R3	3.1600	6.6727	48.61	52.21	6.93	447.42	146.34
R4	3.1659	6.6792	56.79	60.66	7.35	587.23	127.34
R5	3.1585	6.6714	48.22	50.85	7.29	796.74	126.30
R6	3.1585	6.6714	48.22	50.85	7.29	796.74	126.30
R7	3.1626	6.6783	54.22	64.46	8.48	374.94	140.78
R8	3.1597	6.6732	48.92	58.82	6.53	608.12	141.10
R9	4.5103	8.4887	325.67	337.18	7.56	669.21	126.36
R10	4.5092	8.4820	336.90	349.33	6.94	724.90	129.88
R11	4.5022	8.4821	331.77	345.40	5.73	688.28	136.70

$$R = \frac{\sum_{i=1}^{n}\left(\mathrm{PL}_{i,\text{measured}} - \mathrm{PL}_{\text{measured, mean}}\right)^2 - \sum_{i=1}^{n}\left(\mathrm{PL}_{i,\text{predicted}} - \mathrm{PL}_{i,\text{measured}}\right)^2}{\sum_{i=1}^{n}\left(\mathrm{PL}_{i,\text{measured}} - \mathrm{PL}_{\text{measured, mean}}\right)^2} \tag{7}$$

where μ = the mean prediction error in decibel; PL = path loss.

For each hidden neuron instance, average values of the five performance metrics across the ten iterations were computed.

3. Results and discussion

Terrain profile data and path loss data obtained through drive test measurements are presented and analyzed in this section. Also, the performance results of the ANN models that were obtained in the process of experimentation are interpreted and discussed. Furthermore, an optimal ANN model was identified based on prediction accuracy and generalization ability. Finally, a comparative analysis of the prediction results of the developed ANN model and those of Hata, COST 231, ECC-33, and Egli was performed to validate the choice of feed-forward network approach as the optimal option for path loss predictions.

In this study, the propagation environments are described in terms of their respective terrain profile. Table 1 presents the terrain profile characteristics of the survey routes (R1-R11). It can be seen that the average values of latitude, longitude, elevation, altitude, clutter height, and distance from the serving base station transmitter differ essentially from one route to another. Also, the corresponding mean path loss varies across the eleven routes. This occurs because the radio propagation mechanism in wireless channels depends on the terrain profile characteristics along the path of the radio wave transmission (Rappaport, 1996). Figure 2(a)–(n) show the boxplots of latitude, longitude, elevation, altitude, clutter height, and distance in the training and testing datasets, respectively. The statistical distribution of the terrain profile data, and their corresponding path loss data, further proved the differences between the routes under investigation.

Data obtained from extensive field measurement campaigns along the eleven routes in Ota and Ilorin, Nigeria were combined together and pre-processed. The resulting data collected from real scenarios along the eleven survey routes is considered large enough to prevent any case of overfitting during the ANN model development. The comprehensive data-set was divided into training

Figure 2 . Boxplots of variables in (a)–(g) training data-set (h)–(n) testing data-set.

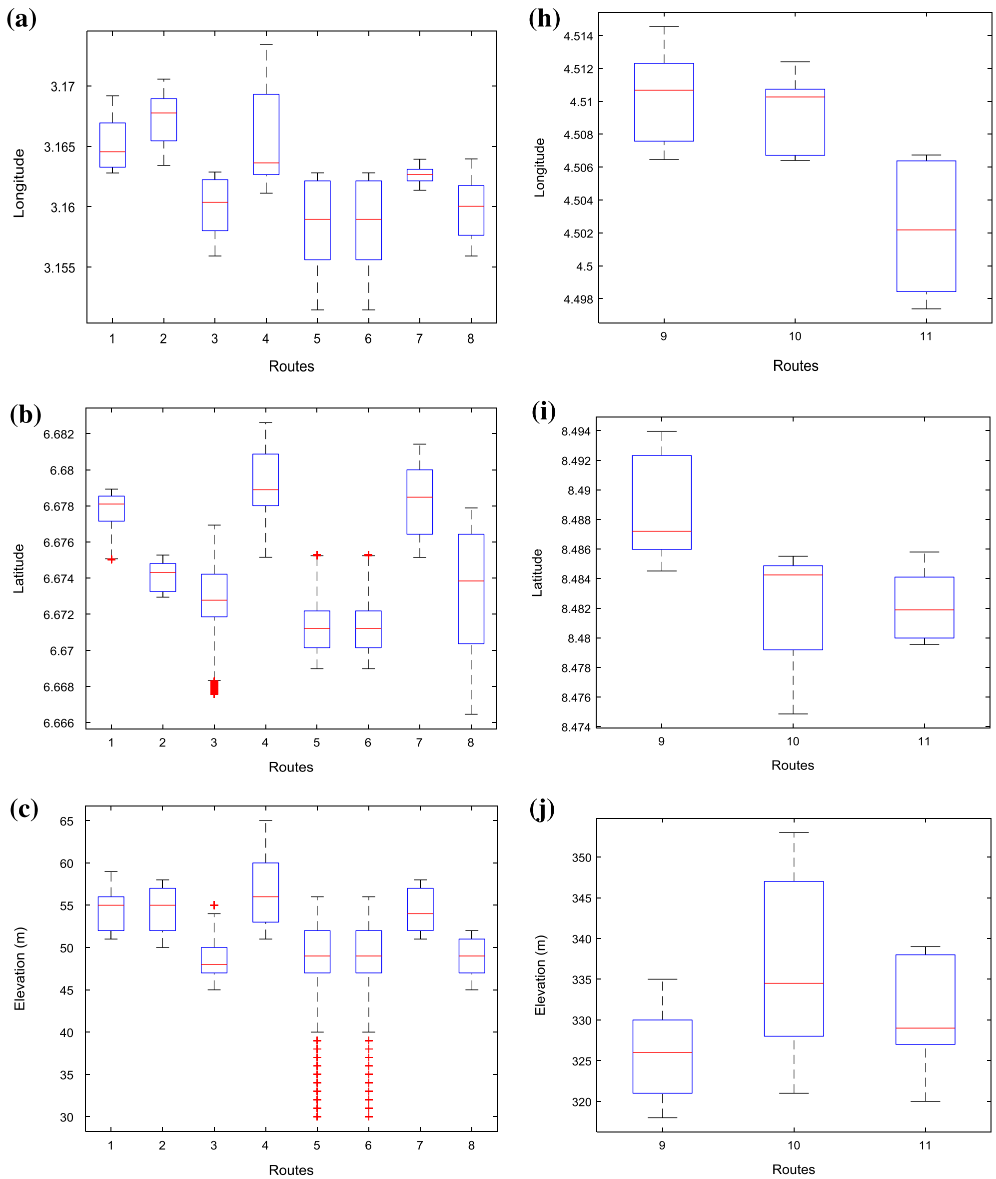


Figure 2. (*Continued*).

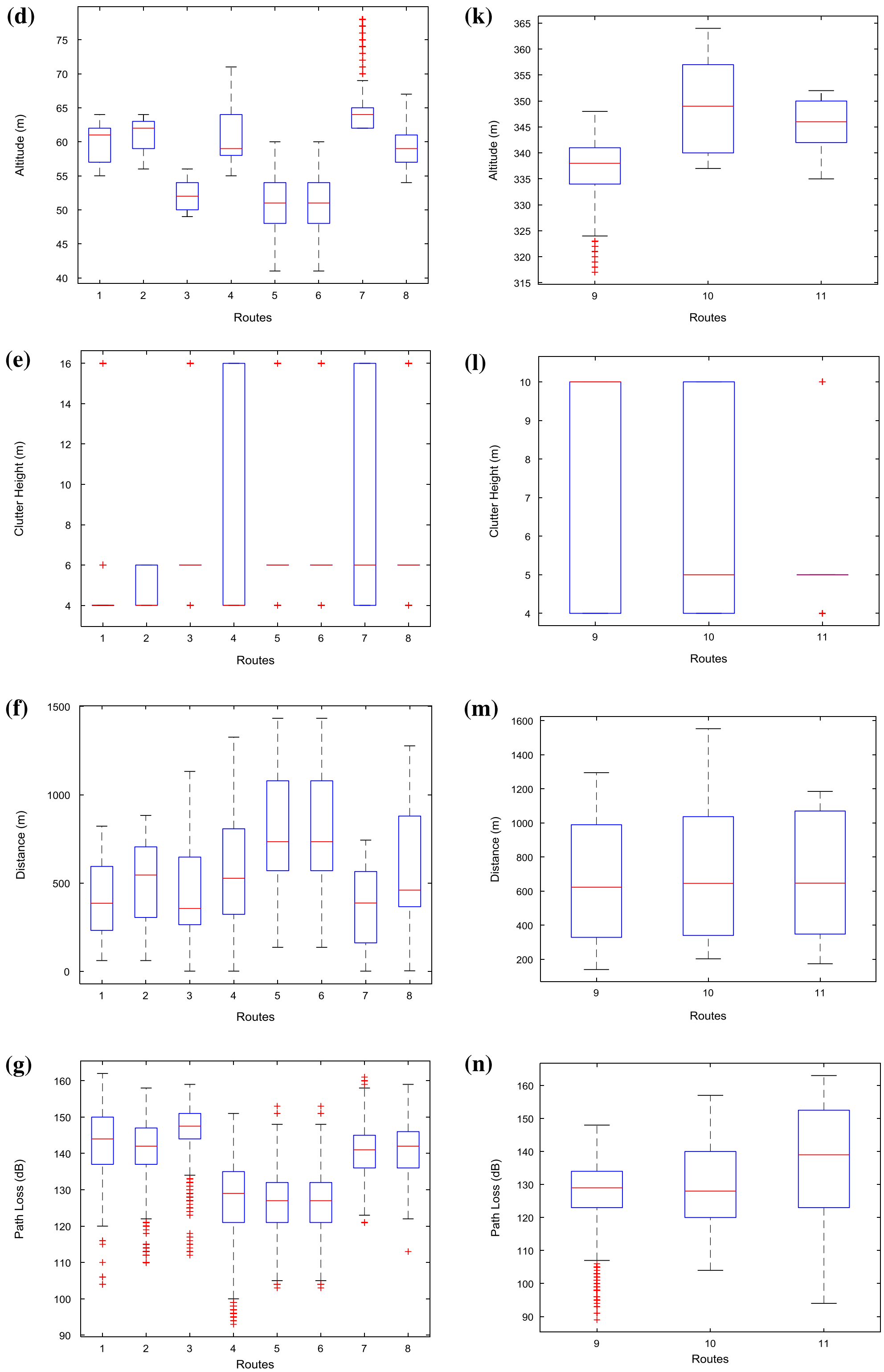
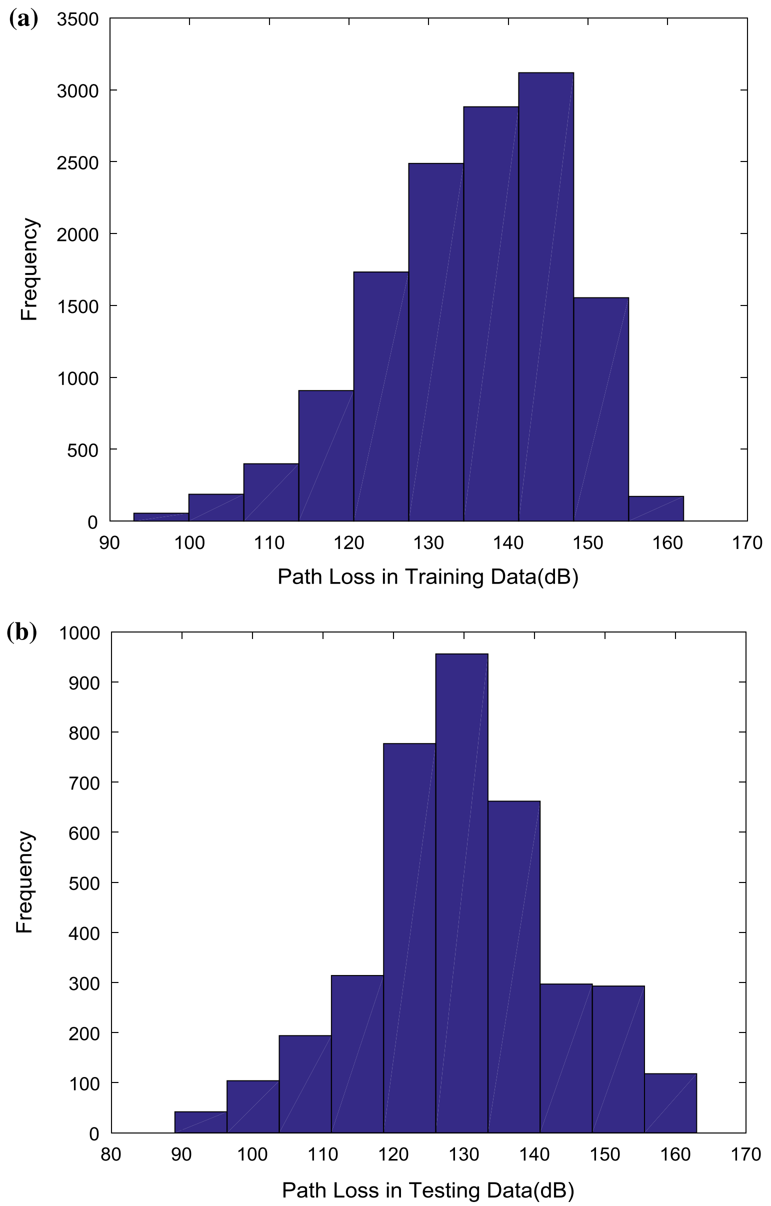


Figure 3. Frequency distributions of path loss in (a) training data and (b) testing data.



and testing sub-datasets for model training and model generalization ability testing, respectively. Tables 2 and 3 present the first-order descriptive statistics (mean, median, mode, standard deviation, variance, kurtosis, skewness, range, minimum, and maximum) of the input and output variables in the training and testing datasets, respectively. The statistical distributions show that the instances of the testing data are different from those of the training data. These facts are presented to ensure that the generalization ability test conducted for the developed ANN model is valid. The claim is further substantiated with the frequency distribution plots of path loss in the training and testing datasets, as shown in Figure 3(a) and (b), respectively.

Single-layered FFNNs were trained with the training data-set based on the Levenberg–Marquardt learning algorithm. The number of neurons in the hidden layer was varied from 1 to 50 to determine the ANN model with the best prediction accuracy and generalization ability. The prediction error is plotted against the number of hidden neurons in Figure 4. There seem to be no obvious changing rules that govern the variation in the prediction error as the number of hidden neurons increases. However, it was observed that the correlation between the target path loss values and the predicted path loss values increases as the number of hidden neurons increases. The response of *R* to changes in the number of hidden neurons is illustrated in Figure 5.

The optimal ANN model was determined based on MAE, MSE, RMSE, standard deviation, and *R* of the predicted path loss values relative to the target path loss values in both training and testing datasets. Results of the machine learning processes show that the FNN architecture that employs the tangent activation function and 48 hidden neurons produced the least prediction error with MAE,

Figure 4. Relationship between the number of hidden neurons and RMSE.

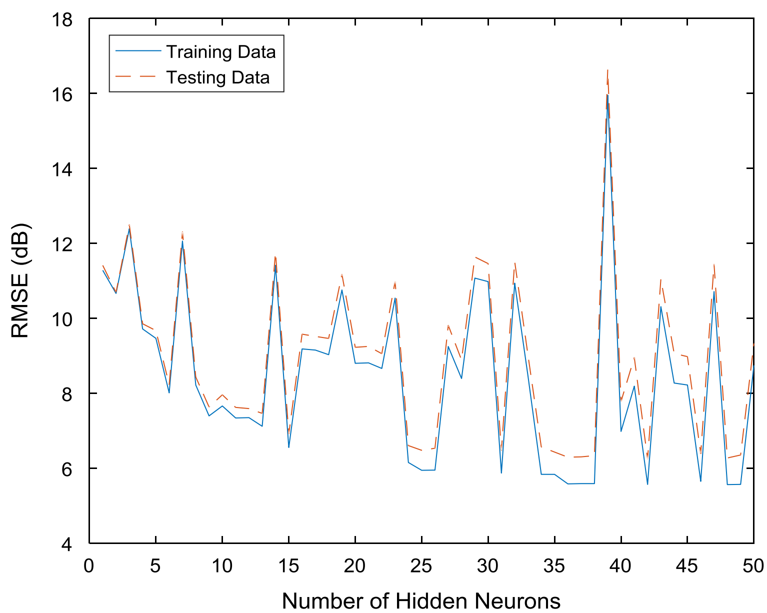


Figure 5. Relationship between the number of hidden neurons and *R*.

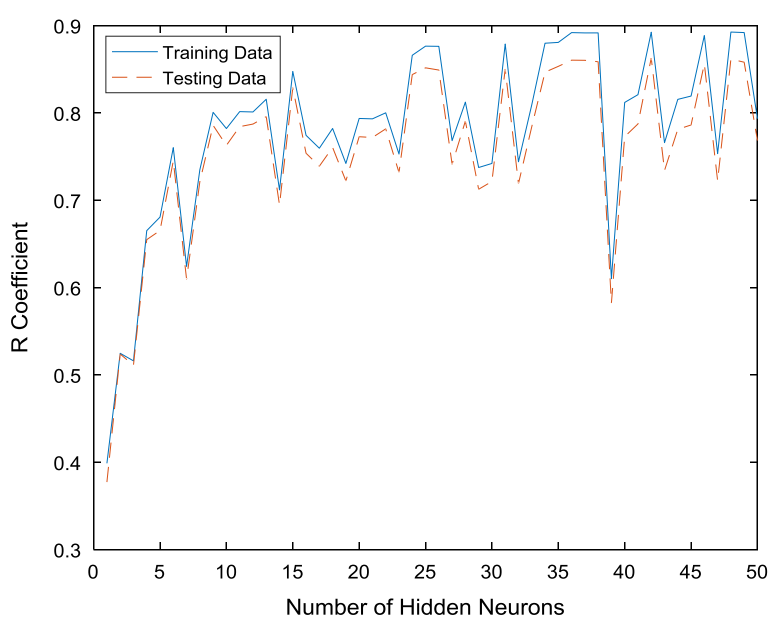


Table 2. Descriptive statistics of input and output variables in training data-set

	Longitude	Latitude	Elevation (m)	Altitude (m)	Clutter height (m)	Distance (m)	Path loss (dB)
Mean	3.3525	6.9310	91.09	97.62	6.86	586.65	135.19
Median	3.1626	6.6756	52.00	59.00	6.00	539.00	136.00
Mode	3.1629	6.6751	52.00	55.00	6.00	709.00	145.00
Standard deviation	0.4690	0.6309	96.97	99.32	3.79	354.04	12.34
Variance	0.2200	0.3981	9,402.79	9,863.83	14.35	125,341.53	152.39
Kurtosis	5.2368	5.2371	5.24	5.24	4.46	2.33	3.05
Skewness	2.0582	2.0583	2.05	2.05	1.68	0.46	−0.57
Range	1.3630	1.8275	323.00	323.00	12.00	1,552.00	74.00
Minimum	3.1515	6.6665	30.00	41.00	4.00	1.00	89.00
Maximum	4.5145	8.4940	353.00	364.00	16.00	1,553.00	163.00
Total samples	9,074	9,074	9,074	9,074	9,074	9,074	9,074

Table 3. Descriptive statistics of input and output variables in testing data-set

	Longitude	Latitude	Elevation (m)	Altitude (m)	Clutter height (m)	Distance (m)	Path loss (dB)
Mean	3.3581	6.9388	92.09	98.63	6.94	586.58	134.79
Median	3.1626	6.6756	52.00	59.00	6.00	539.00	136.00
Mode	3.1574	6.6700	52.00	58.00	6.00	356.00	145.00
Standard deviation	0.4752	0.6392	98.31	100.66	3.90	354.46	12.33
Variance	0.2258	0.4085	9,665.56	10,131.94	15.22	125,642.90	152.15
Kurtosis	5.0327	5.0330	5.04	5.03	4.13	2.33	2.88
Skewness	2.0080	2.0081	2.00	2.00	1.60	0.46	−0.48
Range	1.3631	1.8274	322.00	323.00	12.00	1,541.00	69.00
Minimum	3.1515	6.6665	30.00	41.00	4.00	1.00	94.00
Maximum	4.5145	8.4939	352.00	364.00	16.00	1,542.00	163.00
Samples	1,297	1,297	1,297	1,297	1,297	1,297	1,297

MSE, RMSE, standard deviation, and R values of 4.21 dB, 30.99 dB, 5.56 dB, 5.56 dB, and 0.89, respectively. Regarding the generalization ability, the predictions of the optimal ANN model yielded MAE, MSE, RMSE, standard deviation, and R values of 4.74 dB, 39.38 dB, 6.27 dB, 6.27 dB, and 0.86, respectively when tested with new data not previously included in the training process. The network architecture of the optimal ANN model is shown in Figure 6.

As mentioned earlier, the complete training data-set was divided into 70% training, 15% validation, and 15% testing sub-datasets. Figure 7 shows the degree of correlation between the path loss

Figure 6. Optimal ANN model architecture.

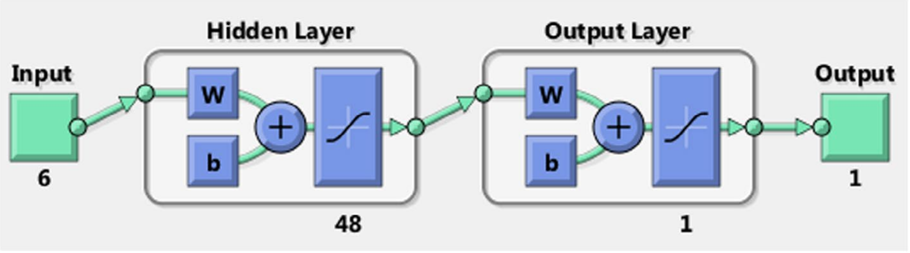
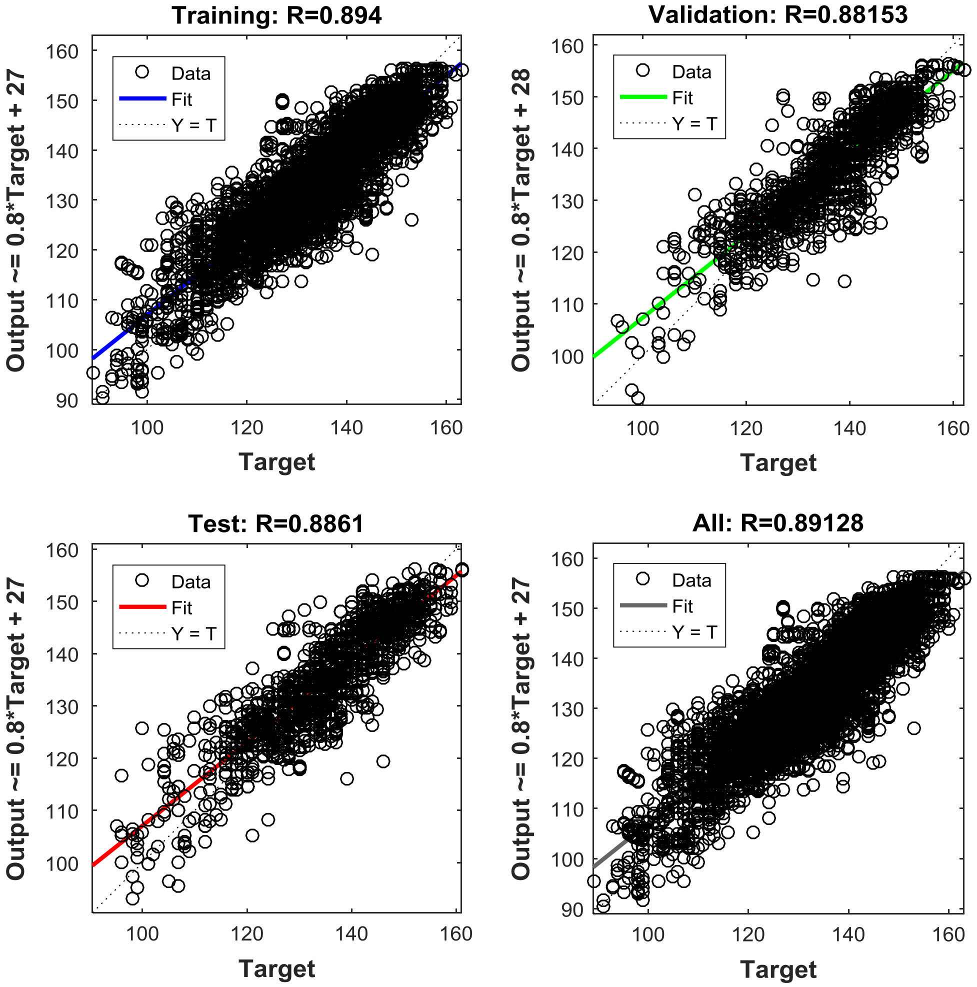


Figure 7. *R* values for optimal ANN model training and validation.



values predicted by the optimal ANN model, and their corresponding target path loss values in the three sub-datasets of the training data-set. *R* values of 0.8940, 0.8815, and 0.8861 were obtained for training, validation, and testing, respectively. The overall *R* value for the model training process is 0.8913. These values are high enough to guarantee high prediction accuracy during radio network planning and optimization within Nigerian propagation environments. For the testing data, the degree of correlation between the predicted output values and the target values is depicted in Figure 8. With an *R* value of 0.8599 for the testing data-set, the developed optimal ANN model proved to have a good generalization ability, considering instances that are not included previously in the model training process.

Figure 8. *R* value for optimal ANN model testing.

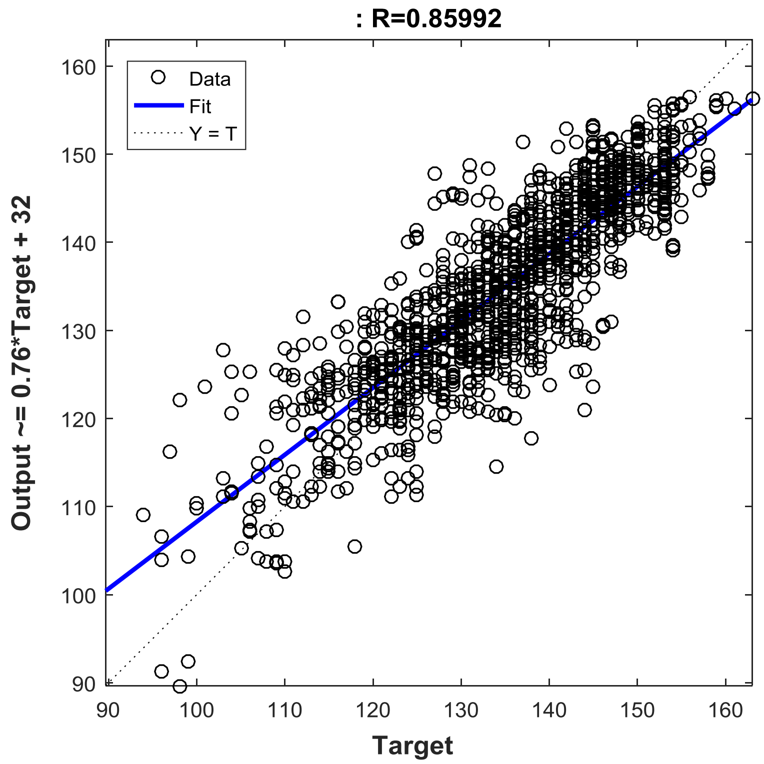
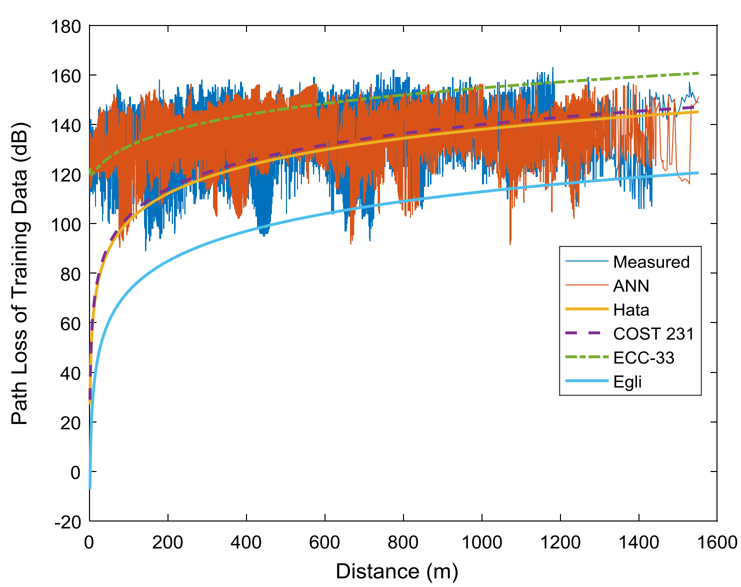
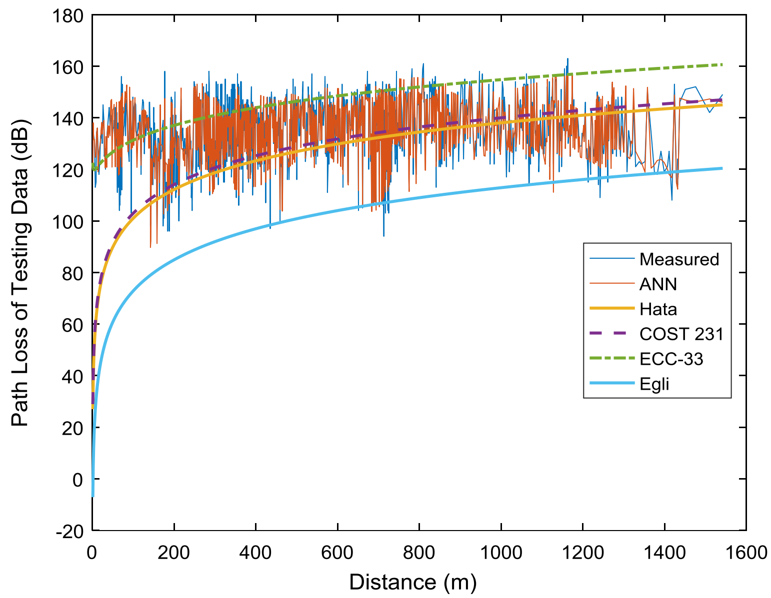


Figure 9. Path loss predictions of ANN and empirical models on training data.



Finally, a comparative analysis of the prediction results of the developed ANN model, and those of Hata, COST 231, ECC-33, and Egli, was performed to validate the choice of the feed-forward network approach as the optimal option for path loss predictions. The measured path loss values and the corresponding predicted path loss values were plotted against the distance of the mobile station from the base station transmitter in Figures 9 and 10. The prediction accuracy of the ANN model was compared to those of the empirical models (Hata, COST 231, ECC-33, and Egli) using the training and testing datasets. In general, it was observed that the four empirical models present the same overall behaviour. However, the ECC-33 model over-predicted the path loss, while the Egli model under-predicted the path loss across the distance range considered. On the other hand, the Hata and COST

Figure 10. Path loss predictions of ANN and empirical models on testing data.



231 models were more conservative, but the prediction results of the ANN model were more accurate. As can be observed, the results achieved for both training and testing datasets using the ANN now present a much greater resemblance with the measured data. In fact, the ANN results resemble a smoothed version of the original data, accurately mimicking the fluctuations detected at different distances.

The results of the performance evaluation are presented in Tables 4 and 5 for the training and testing datasets, respectively. The goodness of the results achieved via the ANN model can be further emphasized by observing the results shown in Tables 4 and 5. Overall, it can be seen that empirical models are unable to accurately represent the measurement data in the 1,800 MHz band for the target regions of Nigeria. In particular, we consider that such failure is associated to the typical environmental conditions, including a combination of buildings and dense foliage, which make them significantly different from the conditions used to tune these models. For the training data, findings showed that the ANN model achieved average values of 4.21 dB, 30.99 dB, 5.56 dB, 5.56 dB and 0.89,

Table 4. Performance evaluation of ANN and empirical models on training data

Model	MAE (dB)	MSE (dB)	RMSE (dB)	Standard deviation (dB)	*R*
Hata	15.43	399.67	19.99	17.28	0.16
COST 231	14.63	364.32	19.09	17.28	0.16
ECC-33	13.85	307.22	17.53	13.74	0.16
Egli	36.38	1,642.57	40.53	18.11	0.16
ANN	4.21	30.99	5.56	5.56	0.89

Table 5. Performance evaluation of ANN and empirical models on testing data

Model	MAE (dB)	MSE (dB)	RMSE (dB)	Standard deviation (dB)	*R*
Hata	15.46	402.50	20.06	17.57	0.14
COST 231	14.71	368.54	19.20	17.57	0.14
ECC-33	14.05	318.51	17.85	13.84	0.15
Egli	36.01	1,628.11	40.35	18.42	0.14
ANN	4.74	39.38	6.27	6.27	0.86

for the MAE, MSE, RMSE, standard deviation, and *R* performance metrics, respectively. On the other hand, for the testing data, it was observed that the ANN model achieved mean values of 4.74 dB, 39.38 dB, 6.27 dB, 6.27 dB and 0.86, for the MAE, MSE, RMSE, standard deviation, and *R* performance metrics, respectively. The prediction errors were relatively high for the testing data because it contained data instances that were not previously included in the training process. Nevertheless, these values represent a dramatic difference compared to the results shown in Tables 4 and 5 for empirical models, highlighting that a satisfactory prediction accuracy and good generalization ability is now achieved using the FFNN approach.

4. Conclusion

To achieve the vision of a ubiquitous Internet, full mobile network coverage stands as a basic requirement. However, many locations worldwide still do not benefit from adequate cellular network coverage, meaning that more efforts are required to improve cellular network planning in order to achieve a successful network deployment.

Having the strategic goal of achieving digital inclusion in mind, the contribution of this paper focused on developing an optimal model for path loss during radio network planning and optimization based on a FFNN technique. Drive test measurements were carried out in Canaanland Ota, Ogun State, Nigeria and Ilorin, Kwara State, Nigeria, to obtain path loss data at varying distances from eleven different 1,800 MHz base station transmitters. The effectiveness of traditional empirical models (Hata, COST-231, ECC-33, and Egli) for path loss predictions in these Nigerian propagation environments was evaluated. Experimental data showed that these models failed to correctly predict path loss values at different distances, especially at low distances, where they tend to be overly optimistic. To solve the aforementioned challenge, Single-layered FFNNs were trained with normalized terrain profile data (longitude, latitude, elevation, altitude, clutter height) and normalized distances to produce the corresponding path loss values based on the Levenberg–Marquardt algorithm. The number of neurons in the hidden layer was varied (1–50) to determine the Artificial Neural Network (ANN) model with the best prediction accuracy.

The performances of the ANN models were evaluated based on the Mean Absolute error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), standard deviation, and regression coefficient (R) metrics. Results of the machine learning processes showed that the FNN architecture that employs a tangent activation function and 48 hidden neurons produced the least prediction error. For the training data-set, experimental results showed that performance metrics experienced a substantial improvement: MAE decreased from 36.38 dB to 4.21 dB; MSE decreased from 1,642.57 to 30.99 dB; RMSE decreased from 40.53 to 5.56 dB; standard deviation decreased from 18.11 to 5.56 dB; and *R* increased from 0.16 to 0.89. Concerning the generalization ability, the evaluation of the predictions produced by the optimal ANN model also showed significant improvements: MAE decreased from 36.01 to 4.74 dB; MSE decreased from 1,628.1 to 39.38 dB; RMSE decreased from 40.35 to 6.27 dB; standard deviation decreased from 18.42 to 6.27 dB; and *R* increased from 0.14 to 0.86, when tested with new data not previously included in the training process.

In summary, the performance of the developed ANN model proved to be optimal in terms of prediction accuracy and generalization ability when compared to those of widely used empirical models (Hata, COST 231, ECC-33, and Egli). As future work, we plan to collect more data from other propagation environments and across other frequency bands for a robust path loss prediction model development. Also, other machine learning approaches shall be exploited to further improve the prediction accuracy and generalization ability of path loss models for efficient radio network planning and optimization in Nigeria.

Acknowledgements
This work is carried out under the IoT-Enabled Smart and Connected Communities (*SmartCU*) Research Cluster of Covenant University. The authors wish to appreciate the Center for Research, Innovation, and Discovery (CUCRID) of Covenant University, Ota, Nigeria for the partial funding of the research.

Funding
This work was supported by Covenant University [grant number CUCRID-SMARTCU-000343].

Author details
Segun I. Popoola[1]
E-mails: segun.popoola@covenantuniversity.edu.ng, segunpopoola15@gmail.com
ORCID ID: http://orcid.org/0000-0002-3941-5903
Emmanuel Adetiba[1,2]
E-mail: emmanuel.adetiba@covenantuniversity.edu.ng
ORCID ID: http://orcid.org/0000-0001-9227-7389
Aderemi A. Atayero[1]
E-mail: atayero@covenantuniversity.edu.ng
ORCID ID: http://orcid.org/0000-0002-4427-2679
Nasir Faruk[3]
E-mail: faruk.n@unilorin.edu.ng
ORCID ID: http://orcid.org/0000-0003-4433-8348
Carlos T. Calafate[4]
E-mail: calafate@disca.upv.es
ORCID ID: http://orcid.org/0000-0001-5729-3041
[1] Department of Electrical and Information Engineering, Covenant University, Ota, Nigeria.
[2] HRA, Institute for Systems Science, Durban University of Technology, P.O. Box 1334, Durban, South Africa.
[3] Department of Telecommunication Science, University of Ilorin, Ilorin, Nigeria.
[4] Department of Computer Engineering, Universitat Politècnica de València (UPV), Valencia, Spain.

Citation information
Cite this article as: Optimal model for path loss predictions using feed-forward neural networks, Segun I. Popoola, Emmanuel Adetiba, Aderemi A. Atayero, Nasir Faruk & Carlos T. Calafate, *Cogent Engineering* (2018), 5: 1444345.

Cover image
Source: Authors.

References

Adetiba, E., Iweanya, V. C., Popoola, S. I., Adetiba, J. N., & Menon, C. (2017). Automated detection of heart defects in athletes based on electrocardiography and artificial neural network. *Cogent Engineering,* 4(1), 1411220.

Adetiba, E., & Olugbara, O. O. (2015). Lung cancer prediction using neural network ensemble with histogram of oriented gradient genomic features. *The Scientific World Journal, 2015*.

Adeyemo, Z. K., Ogunremi, O. K., & Ojedokun, I. A. (2016). Optimization of Okumura-Hata model for long term evolution network deployment in Lagos, Nigeria. *International Journal on Communications Antenna and Propagation (IRECAP), 6*(3), 146–152. doi:10.15866/irecap.v6i3.9012

Akhoondzadeh-Asl, L., & Noori, N. (2007). *Modification and tuning of the universal Okumura-Hata model for radio wave propagation predictions*.

Al Salameh, M. S., & Al-Zu'bi, M. M. (2015). *Prediction of radiowave propagation for wireless cellular networks in Jordan.* Paper presented at the Knowledge and Smart Technology (KST), 2015 7th International Conference on.

Alamoud, M. A., & Schütz, W. (2012). *Okumura-Hata model tuning for TETRA mobile radio networks in Saudi Arabia*.

Armenta, A., Serrano, A., Cabrera, M., & Conte, R. (2012). The new digital divide: The confluence of broadband penetration, sustainable development, technology adoption and community participation. *Information Technology for Development, 18*(4), 345–353. https://doi.org/10.1080/02681102.2011.625925

Begovic, P., Behlilovic, N., & Avdic, E. (2012). *Applicability evaluation of Okumura, Ericsson 9999 and winner propagation models for coverage planning in 3.5 GHZ WiMAX systems*.

Cota, N., Serrador, A., Vieira, P., Beire, A. R., & Rodrigues, A. (2013). *On the use of Okumura-Hat a propagation model on railway communications*.

Erceg, V. (1999). An empirically based path loss model for wireless channels in suburban environments . *IEEE Journal on Selected Areas in Communications,* 1205–1211. https://doi.org/10.1109/49.778178

Farhoud, M., El-Keyi, A., & Sultan, A. (2013). *Empirical correction of the Okumura-Hata model for the 900 MHz band in Egypt*.

Faruk, N., Adediran, Y. A., & Ayeni, A. A. (2013). *Error bounds of empirical path loss models at vhf/uhf bands in kwara state, Nigeria.* Paper presented at the EUROCON, 2013 IEEE.

Faruk, N., Ayeni, A. A., & Adediran, Y. A. (2013). On the study of empirical path loss models for accurate prediction of TV signal for secondary users. *Progress in Electromagnetics Research B, 49*, 155–176. https://doi.org/10.2528/PIERB13011306

GSMA. (2014). The mobile economy 2014. *GSMA Intelligence*.

Hata, M. (1980). Empirical formula for propagation loss in land mobile radio services. *IEEE Transactions on Vehicular Technology, 29*(3), 317–325. https://doi.org/10.1109/T-VT.1980.23859

Hufford, G. A. (1952). An integral equation approach to the problem of wave propagation over an irregular surface. *Quarterly of Applied Mathematics, 9*(4), 391–404. https://doi.org/10.1090/qam/1952-09-04

Ibhaze, A. E., Ajose, S. O., Atayero, A. A.-A., & Idachaba, F. E. (2016). *Developing smart cities through optimal wireless mobile network.* Paper presented at the emerging technologies and innovative business practices for the transformation of societies (EmergiTech), IEEE international conference on.

Luebbers, R. (1984). Propagation prediction for hilly terrain using GTD wedge diffraction. *IEEE Transactions on Antennas and Propagation, 32*(9), 951–955. https://doi.org/10.1109/TAP.1984.1143449

Mardeni, R., & Pey, L. Y. (2010). The optimization of Okumura's model for Code Division Multiple Access (CDMA) system in Malaysia. *European Journal of Scientific Research, 45*(4), 508–528.

Matthews, V. O., Osuoyah, Q., Popoola, S. I., Adetiba, E., & Atayero, A. A. (2017, July 5–7). C-BRIG: A network architecture for real-time information exchange in smart and connected campuses. In *Lecture notes in engineering and computer science: Proceedings of the world congress on engineering 2017* (pp. 398–401). London.

Medeisis, A., & Kajackas, A. (2000). *On the use of the universal Okumura-Hata propagation prediction model in rural areas*. Piscataway, NJ. https://doi.org/10.1109/VETECS.2000.851585

Mohtashami, V., & Shishegar, A. (2012). Modified wavefront decomposition method for fast and accurate ray-tracing simulation. *IET Microwaves, Antennas & Propagation, 6*(3), 295–304. https://doi.org/10.1049/iet-map.2011.0264

Nadir, Z., & Ahmad, M. I. (2010). *Pathloss determination using Okumura-Hata model and cubic regression for missing data for Oman*.

Nimavat, V. D., & Kulkarni, G. (2012). *Simulation and performance evaluation of GSM propagation channel under the urban, suburban and rural environments.* Paper presented at the communication, information & computing technology (ICCICT), 2012 international conference on.

Oseni, O. F., Popoola, S. I., Abolade, R. O., & Adegbola, O. A. (2014). Comparative analysis of received signal strength prediction models for radio network planning of GSM 900 MHz in Ilorin, Nigeria. *International Journal of Innovative Technology and Exploring Engineering,* 4(3), 45–50.

Oseni, O. F., Popoola, S. I., Enumah, H., & Gordian, A. (2014). Radio frequency optimization of mobile networks in Abeokuta, Nigeria for improved quality of service. *International Journal of Research in Engineering and Technology, 3*(8), 174–180.

Phillips, C., Sicker, D., & Grunwald, D. (2013). A survey of wireless path loss prediction and coverage mapping methods. *IEEE Communications Surveys & Tutorials, 15*(1), 255–270. https://doi.org/10.1109/SURV.2012.022412.00172

Popoola, S. I., Atayero, A. A., Badejo, J. A., John, T. M., Odukoya, J. A., & Omole, D. O. (2018). Learning analytics for smart campus: Data on academic performances of engineering undergraduates in Nigerian private university. *Data in Brief, 17*, 76–94. https://doi.org/10.1016/j.dib.2017.12.059

Popoola, S. I., Atayero, A. A., & Faruk, N. (2018). Received signal strength and local terrain profile data for radio network planning and optimization at GSM frequency bands. *Data in Brief, 16*, 972–981. https://doi.org/10.1016/j.dib.2017.12.036

Popoola, S. I., Atayero, A. A., Faruk, N., & Badejo, J. A. (2018). Data on the key performance indicators for quality of service of GSM networks in Nigeria. *Data in Brief, 16*, 914–928. https://doi.org/10.1016/j.dib.2017.12.005

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Adetiba, E., & Matthews, V. O. (2017, July 5–7). *Calibrating the standard path loss model for urban environments using field measurements and geospatial data.* Paper presented at the Lecture notes in engineering and computer science: Proceedings of the world congress on engineering 2017 (pp. 513–518). London.

Popoola, S. I., Atayero, A. A., Faruk, N., Calafate, C. T., Olawoyin, L. A., & Matthews, V. O. (2017). *Standard propagation model tuning for path loss predictions in built-up environments.* Paper presented at the International Conference on Computational Science and Its Applications.

Popoola, S. I., Atayero, A. A., Okanlawon, T. T., Omopariola, B. I., & Takpor, O. A. (2018). Smart campus: Data on energy consumption in an ICT-driven university. *Data in Brief, 16*, 780–793. https://doi.org/10.1016/j.dib.2017.11.091

Popoola, S. I., Badejo, J. A., Ojewande, S. O., & Atayero, A. (2017, October 25–27). Statistical evaluation of quality of service offered by GSM network operators in Nigeria. In *Lecture notes in engineering and computer science: Proceedings of the world congress on engineering and computer science 2017* (pp. 69–73). San Francisco.

Popoola, S. I., Misra, S., & Atayero, A. A. (2018). Outdoor path loss predictions based on extreme learning machine. *Wireless Personal Communications*, 1–20.

Popoola, S. I., & Oseni, O. F. (2014a). Empirical path loss models for GSM network deployment in Makurdi, Nigeria. *International Refereed Journal of Engineering and Science, 3*(6), 85–94.

Popoola, S. I., & Oseni, O. F. (2014b). Performance evaluation of radio propagation models on GSM network in urban area of Lagos, Nigeria. *International Journal of Scientific & Engineering Research, 5*(6), 1212–1217.

Rappaport, T. S. (1996). *Wireless communications: Principles and practice* (Vol. 2). New Jersey: Prentice Hall PTR.

Rath, H. K., Verma, S., Simha, A., & Karandikar, A. (2016). *Path Loss model for Indian terrain-empirical approach.* Paper presented at the communication (NCC), 2016 twenty second national conference on.

Salman, M. A., Popoola, S. I., Faruk, N., Surajudeen-Bakinde, N., Oloyede, A. A., & Olawoyin, L. A. (2017). *Adaptive neuro-fuzzy model for path loss prediction in the VHF band.* Paper presented at the computing networking and informatics (ICCNI), 2017 international conference on.

Schneider, I., Lambrecht, F., & Baier, A. (1996). *Enhancement of the Okumura-Hata propagation model using detailed morphological and building data*. Piscataway, NJ. https://doi.org/10.1109/PIMRC.1996.567508

Sotiroudis, S. P., & Siakavara, K. (2015). Mobile radio propagation path loss prediction using Artificial Neural Networks with optimal input information for urban environments. *AEU-International Journal of Electronics and Communications, 69*(10), 1453–1463. https://doi.org/10.1016/j.aeue.2015.06.014

Verma, R., & Saini, G. (2016). *Statistical tuning of Cost-231 Hata model at 1.8 GHz over dense urban areas of Ghaziabad*.

Zelley, C. A., & Constantinou, C. C. (1999). A three-dimensional parabolic equation applied to VHF/UHF propagation over irregular terrain. *IEEE Transactions on Antennas and Propagation, 47*(10), 1586–1596. https://doi.org/10.1109/8.805904