Radio Propagation Modelling of a Typical Sudan Savanna Belt Rural Terrain using Soft-computing and Empirical Techniques

Deme C. Abraham
Department of Computer Science,
University of Jos, Nigeria.
demeabraham@gmail.com, acdeme2000@yahoo.com

***Abstract*- This paper describes the path loss modeling of a typical Sudan Savanna vegetation belt rural terrain using soft-computing and empirical techniques. The widely used COST 231 Hata model was modified via interpolation, and then compared for path loss prediction accuracy with deep learning network based models, using path loss data derived from field strength measurements recorded at an operating frequency of 1800MHz. Results indicate that while the most accurate of the models is the Generalized Regression Neural Network (GRNN) with an average RMSE value of 3.83dB, the adjusted COST 231 Hata model outperforms the Multi-Layer Perceptron Neural Network (MLP-NN) counterpart.**

Keywords—Path Loss; COST 231 Hata; Generalized Regression Neural Network; Multi-Layer Perceptron Neural Network; Least Squares Approximation.

I. INTRODUCTION

A crucial aspect of the characterization of a wireless link is the determination of path loss. Basically, path loss refers to the difference between the power radiated by a transmitter at one end of a link, and the received power recorded by a receiver at the other end of the link. Path loss across a given terrain significantly depends on the nature of the terrain, atmospheric conditions, radiated power, operating frequency, transmitter height, receiver height, etc. Logically, built-up terrains such as metropolitan environments are characterized by higher path loss compared to suburban and rural terrains.

Empirical modelling is one of the most widely used techniques in path loss estimation across a given terrain. Empirical models are essentially created on the basis of extensive measurements acquired from the terrain in question [1]. These models are not globally applicable due variations in terrain nature. Nevertheless, correction factors are usually introduced into a given model in order to improve path loss estimation accuracy.

Recently, techniques based on the deep learning aspect of artificial intelligence have been used in order to predict path loss with greater accuracy. Deep learning networks have been proven to handle complex non-linear function approximation with a greater accuracy than those techniques which are based on linear regression [2][3].

In this study, the terrain under consideration is a typical Sudan Savanna Belt rural terrain. The Sudan Savanna Vegetation Belt of Nigeria essentially comprises of a mixture of scattered trees and short grasses. The terrain under investigation is a thinly populated rural terrain between the city of Bauchi and Darazo town, comprising of scattered houses and trees mostly below 7 meters.

The COST 231 Hata model is adjusted for improved performance, and then compared for prediction accuracy with deep learning artificial intelligence networks, namely the Generalized Regression Neural Network (GRNN) and the Multi-layer Perceptron Neural Network (MLP-NN). The choice of COST 231 Hata is based on its consideration of frequencies up to 2000MHz since this research considers 1800MHz.

II. THE COST 231 HATA MODEL

As described in [2], the COST 231 Hata Model stems from the Hata Model, taking into consideration a wider range of frequencies (500MHz to 200MHz). The Hata model [4] in turn is an extension of the Okumura Model. The model is suitable for path loss prediction in urban, semi-urban, suburban and rural areas. The model expression is given by (1)

$$PL = 46.3 + 33.9logf - 13.82logh_B - a(h_\mathrm{m}) + (44.9 - 6.55logh_B)logd + C \quad (1)$$

Where,

- PL = Median path loss in Decibels (dB)
- C=0 for medium cities and suburban areas
- C=3 for metropolitan areas
- f = Frequency of Transmission in Megahertz (MHz)(500MHz to 200MHz)
- h_B = Base Station Transmitter height in Meters (30m to 100m)
- d = Distance between transmitter and receiver in Kilometers (km) (up to 20kilometers)

- h_m = Mobile Station Antenna effective height in Meters (m) (1 to 10metres)
- $a(h_m)$ = Mobile station Antenna height correction factor as described in the Hata Model for Urban Areas.
- For urban areas, $a(h_m) = 3.20(\log10(11.75hr))^2 - 4.97$, for f > 400 MHz
  For sub-urban and rural areas, $a(h_R) = (1.1\log(f) - 0.7)h_R - 1.56\log(f) - 0.8$

III. THE GENERALIZED REGRESSION NEURAL NETWORK

As described in [2], the Generalized Regression Neural Network (GRNN), proposed by [5] is a type of Artificial Neural Network (ANN) capable of approximating virtually any function given sufficient data. In contrast to back-propagation neural networks, which may require a large number iterations in order to converge to the desired output, the GR-NN does not require iterative training, and usually requires a fraction of the training samples a back-propagation neural network would need [5]. The GRNN is used to solve a variety of problems such as prediction, control, plant process modeling or general mapping problems [6]. As shown in Fig. 1, the GRNN comprises of four layers:

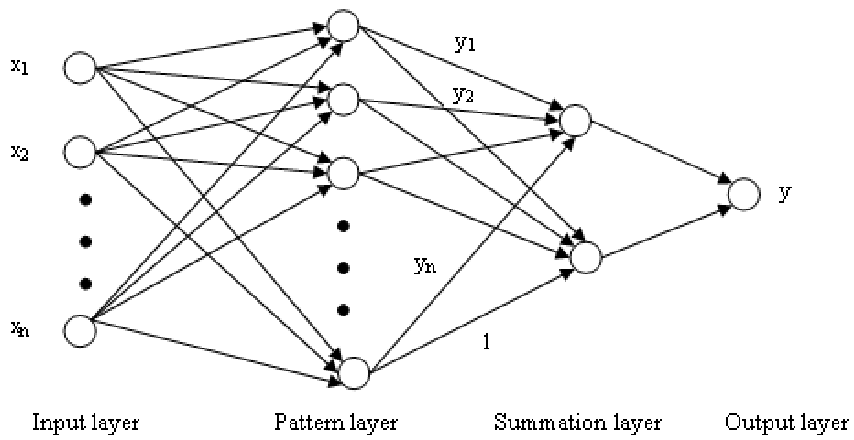

Fig. 1: *Generalized Regression Neural Network Architecture [7]*

Input layer: This is the first layer and it is responsible for sending inputs to the next layer called the pattern layer.

Pattern layer: This layer computes the Euclidean distance between input and training data, and also the activation function.

Summation layer: This layer comprises of two parts: the Numerator and the Denominator. The Numerator sums up products of training data and activation function, while the Denominator sums up activation functions.

Output layer: The single neuron contained in this layer generates the output through division of the Numerator by the Denominator obtained from the previous layer.

The general regression as described by [5] is as follows: given a vector random variable, x, and a scalar random variable, y, and assuming X is a particular measured value of the random variable y, the regression of y on X is given by (2)

$$E[y|X] = \frac{\int_{-\infty}^{\infty} y f(X,y)dy}{\int_{-\infty}^{\infty} f(X,y)dy} \quad (2)$$

If the probability density function $\hat{f}(x,y)$ is unknown, it is estimated from a sample of observations of x and y. The probability estimator $\hat{f}(X,Y)$, given by (3) is based upon sample values X^i and Y^i of the random variables x and y, where n is the number of sample observations and p is the dimension of the vector variable x.

$$\hat{f}(X,Y) = \frac{1}{(2\pi)^{(p+1)/2}\sigma^{(p+1)/n}} \cdot \frac{1}{n}\sum_{i=1}^{n} \exp\left[\frac{(X-X^i)^T(X-X^i)}{2\sigma^2}\right] . \exp\left[\frac{(Y-Y^i)^2}{2\sigma^2}\right] \quad (3)$$

A physical interpretation of the probability estimate $\hat{f}(X,Y)$, is that it assigns a sample probability of width σ (called the spread constant or smoothing factor) for each sample X^i and Y^i, and the probability estimate is the sum of those sample probabilities. The scalar function D_i^2 is given by (4)

$$D_i^2 = (X - X^i)^T(X - X^i) \quad (4)$$

Combining equations (2) and (3) and interchanging the order of integration and summation yields the desired conditional mean $\acute{Y}(X)$, given by (5)

$$\acute{Y}(X) = \frac{\sum_{i=1}^{n} Y^{i} exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \quad (5)$$

The only free network parameter is the smoothing parameter. Neural network training involves finding the optimal value of the smoothing parameter, for which the mean squared error is minimum. As a key advantage over standard feed-forward neural nets, the GNN always converges to a global minimum and hence, has no issues with local minima. It is further stated in [5] that when the smoothing parameter σ is made large, the estimated density is forced to be smooth and in the limit becomes a multivariate Gaussian with covariance σ^2. On the other hand, a smaller value of σ allows the estimated density to assume non-Gaussian shapes, but with the hazard that wild points may have too great an effect on the estimate.

IV. THE MULTI-LAYER PERCEPTRON NEURAL NETWORK

As described in [8], the Multi-Layer Perceptron Neural Network (MLP-NN) is a feed forward neural network trained with the standard back propagation algorithm [9]. They are supervised networks so they require a desired response to be trained. They learn how to transform input data into a desired response, so they are widely used for pattern classification. With one or two hidden layers, they can approximate virtually any input-output map. They have been shown to approximate the performance of optimal statistical classifiers in difficult problems.

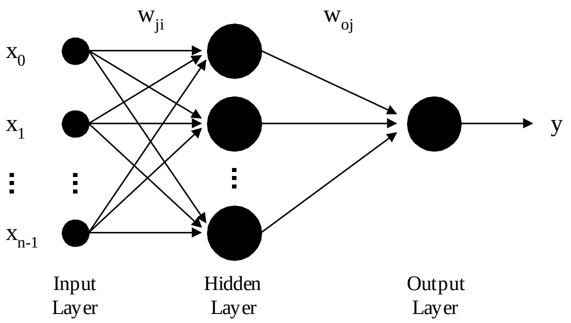

Fig. 2: *Multilayer Perceptron Neural Network with one hidden layer [8]*

As the name implies, a MLP-NN is a network that comprises of an input layer, one or more hidden layers and an output layer. Fig. 2 shows that each neuron of the input layer is connected to each neuron of the hidden layer, and in turn, each neuron of the hidden layer is connected to the single neuron of the output layer. As a result, signal transmission across the entire network can only be in the forward direction, i.e, from the input layer, through the hidden layer and eventually to the output layer. Signals arriving at the inputs propagate forward from neuron to neuron, until they finally arrive at the output neuron and emerge as output signals. Error signals propagate in the opposite direction from the output neuron across the network.

As described in [8] the output of the MLP-NN is describe by (6)

$$y = F_0\left(\sum_{j=0}^{M} w_{0j}\left(F_h\left(\sum_{i=0}^{N} w_{ji} x_i\right)\right)\right) \quad (6)$$

where:

- $\mathbf{w_{oj}}$ represents the synaptic weights from neuron j in the hidden layer to the single output neuron,
- $\mathbf{x_i}$ represents the $\mathbf{i^{th}}$ element of the input vector,
- $\mathbf{F_h}$ and $\mathbf{F_0}$ are the activation function of the neurons from the hidden layer and output layer, respectively,
- $\mathbf{w_{ji}}$ are the connection weights between the neurons of the hidden layer and the inputs.

The learning phase of the network proceeds by adaptively adjusting the free parameters of the system based on the mean squared error E, described in equation (7) between predicted and measured path loss for a set of appropriately selected training examples:

$$E = \frac{1}{2}\sum_{i=1}^{m}(y_i - d_i)^2 \quad (7)$$

where, $\mathbf{y_i}$ is the output value calculated by the network and d_i represents the expected output. When the error between network output and the desired output is minimized, the learning process is terminated and the network can be used in a testing phase with test vectors. At this stage, the neural network is described by the optimal weight configuration, which means that theoretically ensures the output error minimization.

According to [10], a neural network with only one hidden layer can approximate any function with finitely many discontinuities to an arbitrary precision, provided the activation functions of the hidden units are non-linear. Problems that require two or more hidden layers are rarely encountered in practice.

V. MATERIALS AND METHODS

A. Received Power Measurement

The terrain in question is the rural terrain between the towns of Bauchi and Darazo in Northern Nigeria. It is essentially a thinly populated rural area with scattered houses and trees mostly below 7 meters. Received power measurements were recorded from multiple Base Stations scattered across the area. The instrument used was a Cellular Mobile Network Analyser (SAGEM OT 290) capable of measuring signal strength in decibel milliwatts (dBm). The mean isotropic radiated power and transmitter height provided by the mobile network operator are 46dBm and 40 meters respectively. The received power (P_R) readings were recorded at a mobile height of 1.5 meters within the 1800MHz frequency band at intervals of 0.2km after an initial separation of 0.1km.

B. Adjusting the COST231 Hata Model

The COST 231 Hata model was adjusted for improved performance in accordance with the Least Squares Approximation based interpolation technique explicitly described in [11]. By solving the system of normal equations (8) to determine the coefficients a_0,a_1 and a_2, the path loss best fit Least squares function can be formulated as (9)

$$\left.\begin{aligned} \sum_{i=1}^{N} L_i &= Na_0 + a_1\sum_{i=1}^{N} d_i + a_2\sum_{i=1}^{N} {d_i}^2 \\ \sum_{i=1}^{N} d_i\, L_i &= a_0\sum_{i=1}^{N} d_i + a_1\sum_{i=1}^{N} {d_i}^2 + a_2\sum_{i=1}^{N} {d_i}^3 \\ \sum_{i=1}^{N} {d_i}^2\, L_i &= a_0\sum_{i=1}^{N} {d_i}^2 + a_1\sum_{i=1}^{N} {d_i}^3 + a_2\sum_{i=1}^{N} {d_i}^4 \end{aligned}\right\} \quad (8)$$

where,

L_i - path loss values computed from received power measurements

d_i - measurement intervals away from Base Station

N - number of measurements

$$L(d) = a_0 + a_1 d + a_2 d^2 \quad (9)$$

The quotients Q_1, Q_2,. Q_N were obtained at intervals d_1, d_2,. ,d_N respectively, by dividing the Least Squares function value $L(d_i)$, by the COST 231 Hata expression (1) designated as $PL(d_i)$, using (10)

$$Q(d_i) = \frac{L(d_i)}{PL(d_i)} \quad (10)$$

By solving the system of equation (11) to obtain the coefficients b_0 and b_1, the optimal Least Squares function (12) through the quotient points Q_1, Q_2,. Q_N is formulated.

$$\left.\begin{aligned} \sum_{i=1}^{N} Q_i &= Na_0 + a_1\sum_{i=1}^{N} d_i \\ \sum_{i=1}^{N} d_i\, Q_i &= a_0\sum_{i=1}^{N} d_i + a_1\sum_{i=1}^{N} {d_i}^2 \end{aligned}\right\} \quad (11)$$

$$Q(d) = b_0 + b_1 d \quad (12)$$

Finally, the adjusted COST 231 Hata model is obtained by multiplying (1) by (12) to obtain (13).

$$PL_{Adj} = PL \times Q(d) \quad (13)$$

VI. RESULTS AND ANALYSIS

In this study, the path loss prediction models herein considered include the GRNN, MLP-NN, the COST 231 Hata and the adjusted COST 231 Hata. The performance comparison of these models is based on the Root Mean Square Error (RMSE), given by (14), and the Coefficient of Determination (R^2), given by (15). RMSE is essentially a measure of the differences between predicted and observed values. R^2 ranges between 0 and 1, but can be negative, which indicates the model is inappropriate for the data. A value closer to 1 indicates that a greater proportion of variance is accounted for by the model.

$$RMSE = \sqrt{\sum_{i=1}^{N}\frac{(M-P)^2}{N}} \quad (14)$$

Where,

M – Measured Path Loss
P – Predicted Path Loss
N- Number of paired values

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y}_i)^2} \quad (15)$$

Where

y_i is the measured path loss,
$\hat{y}_i$ is the predicted path loss and
$\bar{y}_i$ is the mean of the measured path loss values.

Using the interpolation technique earlier described, the COST 231 Hata was interpolated onto the formulated best fit Least Squares function (16) representing the mean measured path loss values.

$$L(d) = 102.5 + 15.84d - 1.38d^2 \quad (16)$$

The adjustment function obtained is (17).

$$Q(d) = 0.8859 + 0.0049d \quad (17)$$

Hence, the adjusted COST 231 Hata is given by (18), considering C=0 for suburban and rural areas.

$$PL_{Adj} = (0.8859 + 0.0049d) \times (46.3 + 33.9logf - 13.82logh_B - a(h_{\mathrm{m}}) + (44.9 - 6.55logh_B)logd) \quad (18)$$

During neural network training, validation and testing, the set of recorded path loss values for each Base Station was randomly split as follows: training- 50%, validation-10% and 40%-testing. Figs. 3 to 8 depict graphical performance comparisons of the four predictors. It can be clearly observed that the COST

231 Hata overestimates the path loss across all the BTSs, while the others are convergent in performance. The performance indices in Table 1 indicate that the GRNN based predictor with a RMSE value of 3.83dB is the most accurate of all. Furthermore, its R^2 value of 0.86 indicates that it has the best fit. Interestingly, it can be observed that the adjusted COST 231 Hata model with an average RMSE value of 5.62dB is fairly consistent across the BTSs and even outperforms the MLP-NN. This is a testament to the effectiveness of the interpolation technique described in [14]. As stated earlier, the table shows that the COST 231 Hata overestimates the path loss by 16.06dB.

Table 1: Statistical Performance Comparison of Predictors

MODEL	STATS.	BST 1	BST 2	BST 3	BST 4	BST 5	BST 6	MEAN
GRNN	RMSE(dB)	4.47	3.60	3.23	3.27	5.43	2.95	3.83
	R^2	0.86	0.85	0.92	0.93	0.87	0.93	0.89
MLP-NN	RMSE(dB)	5.58	5.23	9.83	8.12	6.32	5.20	6.71
	R^2	0.60	0.69	0.27	0.54	0.82	0.79	0.62
Adjusted COST 231 Hata	RMSE(dB)	8.06	5.32	5.90	5.10	5.25	4.06	5.62
	R^2	0.57	0.87	0.83	0.87	0.89	0.91	0.82
COST 231 Hata	RMSE(dB)	15.51	17.68	17.54	14.49	16.24	14.90	16.06
	R^2	-0.14	-1.05	-0.50	-0.07	-0.10	-0.26	-0.35

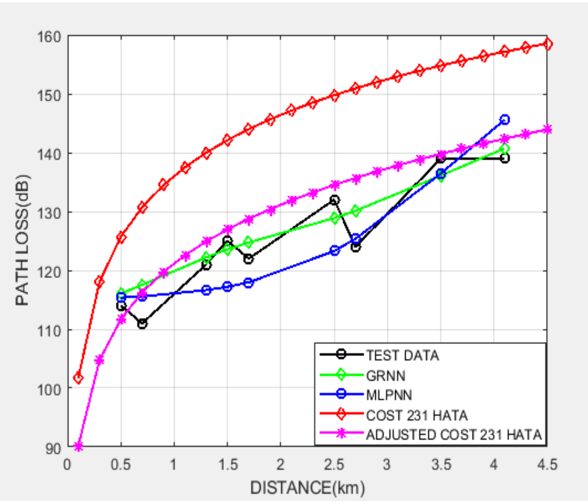

Fig. 3: *BST 1 Model Comparison*

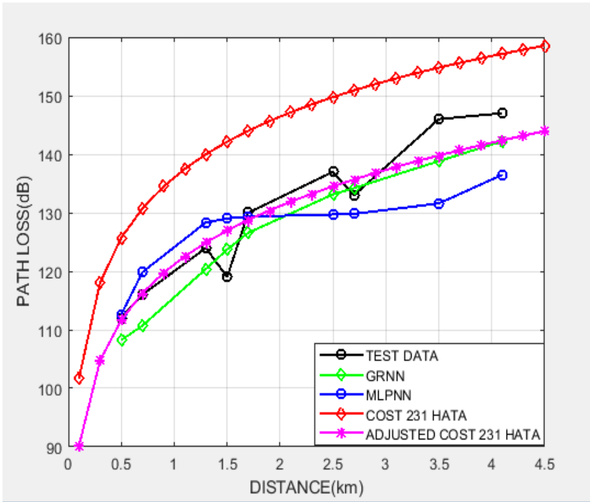

Fig. 4: *BST 2 Model Comparison*

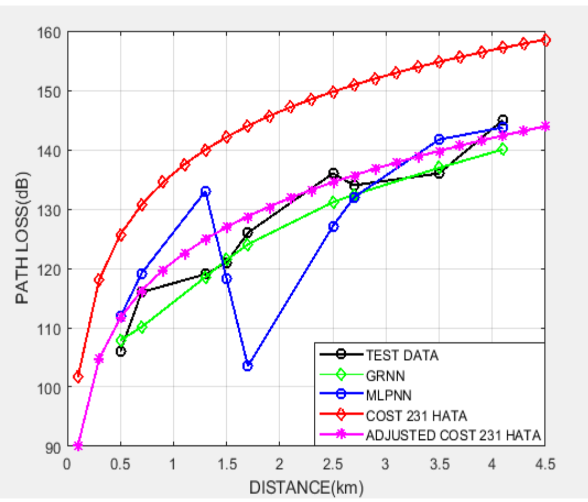

Fig. 5: *BST 3 Model Comparison*

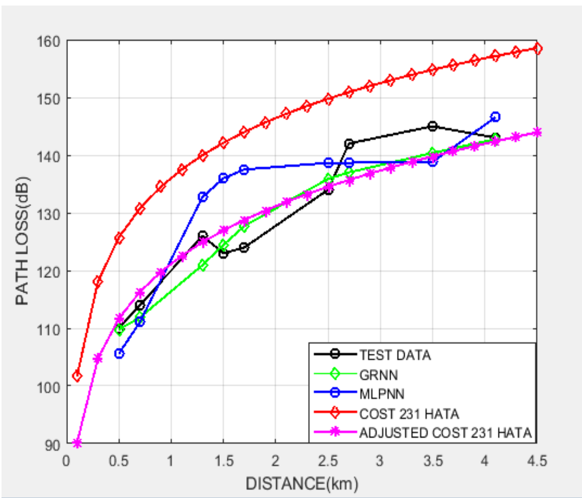

Fig. 6: *BST 4 Model Comparison*

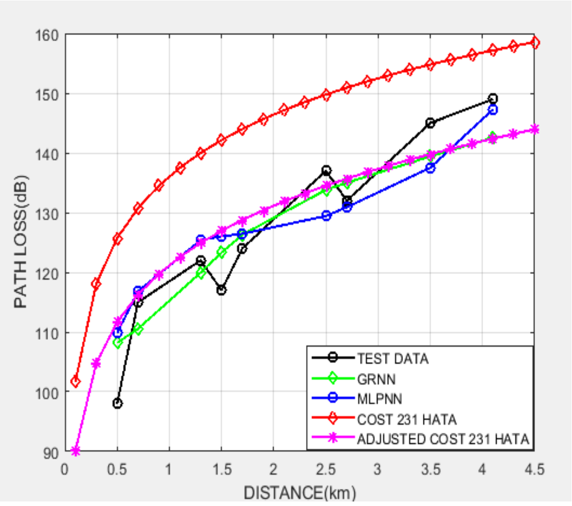

Fig. 7: *BST 5 Model Comparison*

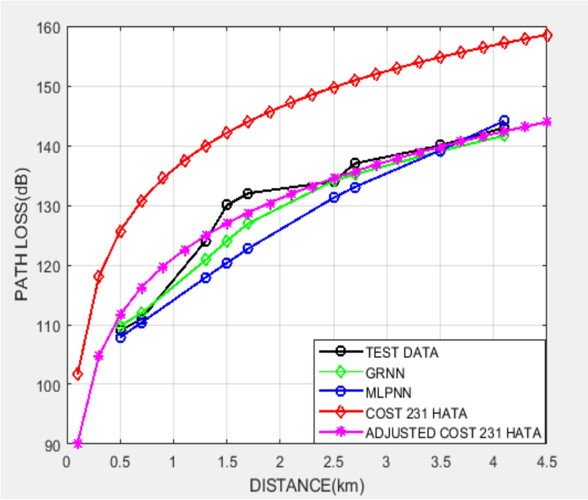

Fig. 8 : *BST 6 Model Comparison*

VII. CONCLUSION

This paper demonstrates that the widely used COST 231 Hata model overestimates the path loss across a typical Sudan Savanna vegetation belt rural terrain when analyzed using data obtained at 1800MHz. However, models based on deep learning networks along with the modified COST 231 Hata offer significant improvements over the actual COST Hata model. Results indicate that the most accurate of the models considered is GRNN with a RMSE value of 3.82dB, while the modified COST 231 Hata outperforms the MLP-NN model.

REFERENCES

[1] Akpado K.A., Oguejiofor O.S., Abe A., Ejiofor A. C. Pathloss Prediction for a typical mobile communication system in Nigeria using empirical models. IRACST – International Journal of Computer Networks and Wireless Communications (IJCNWC), ISSN: 2250-3501 Vol.3, No 2. 2013. Pp. 207-211.

[2] Deme C. A. A Generalized Regression Neural Network Model for Path Loss Prediction at 900 MHz for Jos City, Nigeria. *American Journal of Engineering Research (AJER)*. e-ISSN: 2320-0847 p-ISSN : 2320-0936 Volume-5, Issue-6, pp-01-07 www.ajer.org.

[3] Deme C. A. Development and Comparison of Artificial Neural Network Techniques for Mobile Network Field Strength Prediction across the Jos-Plateau, Nigeria. Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 2458-9403 Vol. 3 Issue 6, June - 2016

[4] Ubom E. A, Idigo, V.E., Azubogu, A.C.O., Ohaneme, C.O., and Alumona, T.L. "Path loss Characterization of Wireless Propagation for South-South Region of Nigeria", International Journal of computer theory and Engineering, 2011, vol.3, N03.

[5] Specht D.F.A. A general regression neural network. *IEEE Transactions on Neural Networks*. 2, 1991, 568-576.

[6] Leszek R. Generalized Regression Neural Networks in Time-Varying Environment*. IEEE Transactions on Neural Networks, Vol. 15, No. 3*, 2004, pp. 576-596.

[7] Sun G., Hoff S. J., Zelle B. C., Nelson M. A. Development and Comparison of Backpropagation and Generalized Regression Neural Network Models to Predict Diurnal and Seasonal Gas and PM_{10} Concentrations and Emissions from Swine Buildings. *American Society of Agricultural and Biological Engineers ISSN 0001-2351. Vol. 51(2):2008,* pp. 685-694

[8] Popescu I., Naforni I., Gavriloaia G. Field Strength Prediction in Indoor Environment with a Neural Network Model:FACTA UNIVERSITATIS (NIS), Series: Electronics and Energetics, 2001. Vol. 14, No. 3, pp 329-336.

[9] Gaurang P., Amit G., Y P Kosta and Devyani P. (2011) "Behaviour Analysis of Multilayer Perceptrons with Multiple Hidden Neurons and Hidden Layers", *International Journal of Computer Theory and Engineering*, Vol. 3, No. 2, pp. 332-337.

[10] Östlin, E., Zepernick, H.-J., & Suzuki, H. Macrocell radio wave propagation prediction using an artificial neural network. IEEE Semiannual Vehicular Technology Conference, 2004, 1, 57–61. 2

[11] Deme C. A., Dajab D. D., Sani M. S. A Discrete Least Squares Approximation Based Algorithm For Empirical Model Adaptation. Journal of Multidisciplinary Engineering Science and Technology (JMEST) ISSN: 2458-9403 Vol. 3 Issue 6, June - 2016