

Path Loss Modeling: A Machine Learning Based Approach Using Support Vector Regression and Radial Basis Function Models

Stephen Ojo^{1*}, Arif Sari², Taiwo P. Ojo³

¹College of Engineering, Anderson University, Anderson, SC, USA

²Department of Management Information Systems, Girne American University, Girne, Cyprus

³Computer Science Department, Florida A&M University, Tallahassee, FL, USA

Email: *stephenajo70@yahoo.com, *sojo@andersonuniversity.edu, arifsari@gau.edu.tr, taiwo1.ojo@famou.edu

How to cite this paper: Ojo, S., Sari, A. and Ojo, T.P. (2022) Path Loss Modeling: A Machine Learning Based Approach Using Support Vector Regression and Radial Basis Function Models. *Open Journal of Applied Sciences*, 12, 990-1010.
<https://doi.org/10.4236/ojapps.2022.126068>

Received: May 28, 2021

Accepted: June 25, 2022

Published: June 28, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Path loss prediction models are vital for accurate signal propagation in wireless channels. Empirical and deterministic models used in path loss predictions have not produced optimal results. In this paper, we introduced machine learning algorithms to path loss predictions because it offers a flexible network architecture and extensive data can be used. We introduced support vector regression (SVR) and radial basis function (RBF) models to path loss predictions in the investigated environments. The SVR model was able to process several input parameters without introducing complexity to the network architecture. The RBF on its part provides a good function approximation. Hyperparameter tuning of the machine learning models was carried out in order to achieve optimal results. The performances of the SVR and RBF models were compared and result validated using the root-mean squared error (RMSE). The two machine learning algorithms were also compared with the Cost-231, SUI, Egli, Freespace, Cost-231 W-I models. The analytical models overpredicted path loss. Overall, the machine learning models predicted path loss with greater accuracy than the empirical models. The SVR model performed best across all the indices with RMSE values of 1.378 dB, 1.4523 dB, 2.1568 dB in rural, suburban and urban settings respectively and should therefore be adopted for signal propagation in the investigated environments and beyond.

Keywords

Support Vector Regression, Radial Basis Function, Machine Learning, Path Loss, Empirical, Deterministic

1. Introduction

In a post-pandemic world, high speed internet access is not a luxury but a necessity because many human activities have gone virtual. High quality of service among wireless subscribers is therefore the lifeline for digital inclusion in a world ravaged by the overall negative effect of the pandemic. To ensure an adequate and quality signal level for users in wireless networks, signal propagation models are useful and essential. The rapid expansion in wireless communication services because of the increase in subscribers' demand has brought path loss predictions to the front burner [1] [2]. Network planning and optimization have occupied a central place in wireless communication because of their pivotal role in signal characterizations. Path loss is the attenuation in signal strength as signals propagate from the transmitter to the receiver in wireless channels. This gradual reduction in the strength of the signal as it propagates in air is the path loss. Path loss on its own is inevitable because the mechanism for electromagnetic propagation like refraction, scattering and diffraction are diverse but can be accurately represented and characterized [3] [4].

Signal propagation models are important for interference analysis, frequency assignment estimation and network planning. The accurate characterization of signal path loss in a wireless channel will help network providers make in-depth estimation before deployment. The essentiality of propagation models for accurate path loss representation is vital for its sustainability and adaptivity [5]. Several propagation models have been developed and used for signal propagation in different frequency bands across different countries and territories. A model developed for a particular propagation environment performed optimally in that environment but failed when deployed to another environment.

This has been a major setback for propagation models that were developed based on empirical measurements. The reliability of the propagation models fades when the deployment outside the initial region of development is carried out [6] [7] [8]. This is because the model parameters are a function of the specific environment. Antenna height, width of buildings, building to building distance and heights of the transmitter and receiver antenna are a function of the environment and therefore become the dominant factors in wireless signal predictions. When a developed model is put to use in other locations, correction factors are always being added to each of the model parameters so as to make them give an optimal solution [9] [10]. This practice of adding certain factors to the model parameters has not provided an accurate solution as the ultimate model after modification ends up overpredicting or underpredicting the network. Propagation models are broadly classified into empirical and deterministic models. Empirical models are based on experimental measurements within a particular environment. These models are developed based on measurements of base station heights, frequency, distance between transmitter and receiver, building to building distance and a host of others [11] [12] [13]. These models are simple and easy to develop but do not guarantee high level accuracy because in a bid to

get them simplified many components are left out. The need to reduce complexity in empirical models in many instances has necessitated the need to use just few environmental parameters in modeling. Empirical models are thus a trade-off between accuracy and simplicity.

Deterministic models are built based on the law governing electromagnetic induction. They are based on several radio wave propagation characteristics in a particular propagation environment [14] [15] [16] [17]. They produced a high level accuracy than the empirical models but introduce so many complexities into the developed model. Such complexities introduced in deterministic models and the lack of accuracy as in the case of empirical models made them not suitable for signal propagation in complex multipath propagation environments. There is need for a model that will give a high level of accuracy and also provides easy representation. The shortcomings of empirical and deterministic models have necessitated the need for path loss models with greater functionality and adaptability [18] [19] [20]. This is the introduction of machine learning algorithms for path loss predictions in wireless channels.

The introduction of machine learning algorithms to signal propagation modeling provides a robust network architecture, robust adaptability and extensive data usability. With machine learning adoption to signal propagation, many components from the transmitter to the receiver can be modelled effectively. Machine learning algorithms will capture all the environmental parameters without necessarily introducing any complexities into the model. Thousands of data sets can be introduced in order to ensure accurate representation and adaptability. The path loss prediction problem is a supervised regression problem in which the main objective is to effectively train the data sets, such that the cross validation can produce a set of optimal weights that can guarantee an accurate model [21] [22].

Extensive dataset is needed for this, because the data will be split into training and test sets in such a way that the model can be cross validated with the test dataset after training in order to ensure the accuracy. Any path loss prediction model developed with machine learning can be effectively adopted for use in any other environment outside of the initial place of development. This is because the model is generic and can be adopted on completely different set of data with an higher accuracy. Function approximation is therefore very easy with ML algorithms by all standards and their accuracies are fore more superior than the empirical and deterministic models using any of the key performance indices. Several ML path loss prediction models have been adopted in signal propagation modeling in wireless environments.

In [23] a neural network model for path loss prediction model was developed in the ultra-wide-band frequency band. The model was trained and the optimal value of the weight was found using backpropagation. The model compared with other empirical models gave the lowest value of RMSE. Also in [24], a multi-layer perceptron neural network path loss prediction model was introduced. The key contribution is that it examined the performances of using several hid-

den layers on the overall network performance. It was discovered that increasing the number of hidden layers in the network increases the predictive accuracy but also introduced some complexities.

Differential evolution was used with artificial neural network in [25]. The model was trained with experimental data in urban environments. The model also employed gradient descent algorithm in conjunction with the backpropagation neural network. The model compared with other experimental studies in the same location gave an more accurate path loss characterization than the other models. A radial basis function path loss model was introduced in [26]. The model was compared with the performance of the multi-layer perceptron neural network and other five existing propagation models. The radial basis function predicted path loss with the highest accuracy and gave the lowest value of error in the considered environments. In [27], an artificial neural network for several frequency bands was developed with experimental data. The performance of the ANN network was examined among the different frequency bands. The ANN network gave the most accurate prediction. The other empirical models over predicted path loss and did not give an accurate signal characterization. Some other applications of machine learning algorithms and methods are given in **Table 1**.

Several other ML algorithms have been implemented in the literature. They ranged from both supervised and unsupervised learning algorithms with the sole

Table 1. Some related works.

Reference	Year	Algorithms and methods	Contributions
[28]	2021	Radial basis function and Multilayer perceptron	Provides accurate path loss prediction using radial basis function and multilayer perceptron algorithms
[29]	2022	Ensemble Machine Learning	Combined several machine learning algorithms to achieve accurate prediction
[30]	2018	Multiplicative calculus	Multiplicative calculus gave predictions with the lowest value of RMSE as compared to the other empirical models
[31]	2022	Atmospheric propagation modeling at microwave frequency	Modelling of path loss with greater accuracy even at microwave frequency
[32]	2022	Prognostic modeling of specific radio attenuation	Thorough prognosis of and modeling of signal attenuation
[33]	2021	Systems dynamic approach with some machine learning methods	Machine learning methods enhanced with blockchain methods
[34]	2021	Automation with machine learning algorithms	Machine learning algorithms for automation of processes
[35]	2022	Machine learning algorithms for EEG signals	MLP algorithms, decision trees and random forest used in signal modeling
[36]	2022	Multilayer perceptron neural network for path loss prediction	Accurate signal prediction and characterization

objective of giving accuracy to path loss modeling. Some Machine learning algorithms have not been explored and even when they are explored, there is need to compare the performances of the different machine learning algorithms for the purpose of generalization.

The contributions of this paper are specified as follows.

- In other to achieve accurate path loss prediction in wireless networks, Support vector regression (SVR) and Radial basis function model were developed with extensive data sets. These machine learning based models were compared to the predictions given by empirical models.
- Several hyperparameters of the RBF model were tuned in order to achieve optimization for accurate system design of the SVR and RBF machine learning algorithms.
- The performances of the SVR and RBF machine learning were compared with five other empirical models for accurate results validation using RMSE.

The remaining part of this work is organized as follows; Section 2 presents the measurement campaign scenario and data pre-processing, the methodology and model developments are given in Section 3. The results are presented in Section 4 and conclusion is contained in Section 5.

2. Measurement Campaign Scenario and Data Preprocessing

Extensive field measurements were collected across base stations in Lefkosia and Kyrenia both located in the Northern part of Cyprus with the aid of drive test. The five environmental features used as input to the machine learning models are elevation, clutter heights, distance between the transmitter and receiving antenna, altitude, building to building distance and the street orientation angle. The received signal strength indicator (RSSI), reference signal received power (RSRP) and the reference signal received quality (RSRQ) were all measured during the drive test which was carried out at 40 km/hr to minimize the effect of Doppler effect. For the drive test, the measurement equipment which are TEMs mobile system dongle software, GPS, 3G mobile phone are connected to a laptop and housed inside a moving vehicle. The height of the base transceiver station for the urban centers where field measurements were collected is 25 m while that of suburban centers is 35 m. The mobile antenna height used is 1.5 m. The reference distance is 100 m and data was measured spanning a transmitter-receiver distance of 3 km. As the mobile station moves away from the transmitting station, the RSRP was measured at each distance. At each distance, 10 readings were recorded and the average was taken in order to ensure data accuracy. The RSRP was then subtracted from the EIRP at each distance to determine the path loss at that point. **Figure 1** provides the field measurement interface during the drive test.

The cross validation was done by dividing the dataset into two as follows. 75% of the dataset was used for training and 25% was used as the test dataset. This was done so that the model can be tested with the data outside of the training sample.

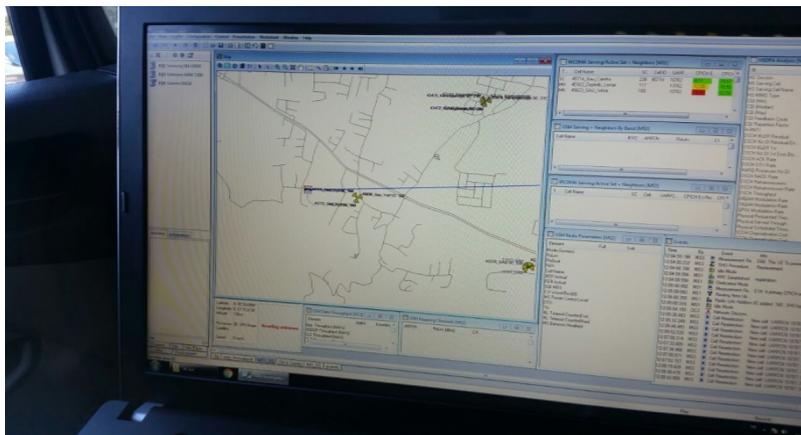


Figure 1. Field measurements during drive test.

The data preprocessing was done and some data normalization was carried out on the measured data. The MinMax scaling normalization was carried out on the data as given in Equation (1)

$$X_{sc} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (1)$$

Path loss is computed as given in Equation (2).

$$\text{Path loss (dB)} = \text{Effective Isotropic Radiated Power} - \text{Received power} \quad (2)$$

$$\text{EIRP} = \rho_1 + \rho_2 + \rho_3 - n_1 - n_2 - n_3 \quad (3)$$

where, ρ_1 is the power of the transmitter in dBm, ρ_2 is the gain of the transmitting antenna, ρ_3 is the gain of the receiving antenna, n_1, n_2, n_3 represents the feeder cable loss, antenna loss, and antenna filter loss, respectively.

The figurative descriptive statistics of the input parameters to the machine models are presented in **Figure 2**. It also represents correlation of each of the input parameters to the output path loss.

3. Methodology and Model Developments

The methodology and model developments are introduced in this section. Two machine learning based models namely; support vector regression and radial basis functions were introduced. The development of this model will be adapted to path loss prediction as in a supervised learning regression problem. Training of both models will be carried out with the same datasets so as to accurately make a justifiable comparison between the performances of both models.

3.1. Support Vector Regression Path Loss Prediction Model

One of the most significant solutions of engineering problems is to be able to approximate a function with another function. Support vector regression is a generalization of the support vector machines which can be essentially used in providing accurate characterization and solution for path loss prediction problems which are typically non-linear regression problems. SVR has the ability to

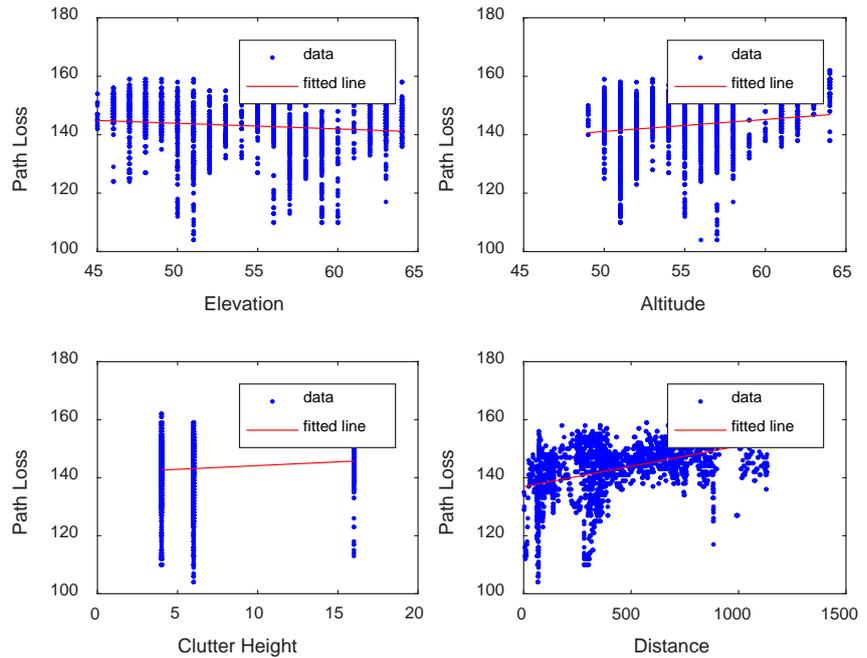


Figure 2. Relationship between each of the input parameters and the output path loss.

transform data which are non-linear separable into a linearly separable form by taking the sets of data initially in the single-dimensional space into an higher dimensional space. After this transformation to the multidimensional space, they become linearly separable. In the high-dimensional place, the functions search for a hyperplane that enables all the datapoints to fall accurately on the hyperplane, with the samples falling on the hyperplane explicitly. The predictive accuracy of the proposed SVR model is due to its capacity to optimize the margin, absence of local minima and the sparseness of the given solution. Thus it is very suitable for function approximation problems.

For the model development, we employed a supervised learning ML algorithm with M training data set. The supervised machine learning method has an input data set mapped correspondingly with the output (target values).

The training data set $M = \{(x_i, y_i) \in R^n * R, i = 1, 2, \dots, K\}$ with K pairs $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$, where the inputs into the SVR model are n dimensional features in vector form $x_i \in R^n$, the output path loss $y_i \in R$ are continuous with K taking as the number of samples in the dataset. The specific hyperplane is expressed in Equation (4) as

$$f(x) = w^T \varnothing(x) + b \quad (4)$$

where x is the input vector and represents all the five input parameters namely, elevation, clutter heights, distance between the transmitter and the receiving antenna, altitude and building to building distance.

$\varnothing(x)$ represents the non-linear mapping function.

b stands for the bias and w is the normal vector that controls and specifies the direction of the hyperplane in the multi-dimensional space.

The predicted path loss from the SVR model is given in Equation (5) as

$$y = \sum_{i=1}^K (c_i - c_i^*) P(x_i, x) + b \quad (5)$$

where, y represents the output predicted path loss from the SVR model, b stands for the bias,

c_i and c_i^* represents the Lagrange multipliers,

x_i represents the input vector which in this case are the five input parameters,

$P(\cdot)$ is represents the kernel function employed in the model development which performs the transformation of the non-linear function from the low-dimensional space to the multi-dimensional space.

K represents the number of sample data for the training without including the test datasets.

The overall performance of the SVR model depends on the type of kernel function used in the development. There are basically three types of kernel function namely; polynomial, Gaussian and Laplacian kernel functions. For this work, we employed the Gaussian kernel function (radial basis) because of the superior performance it gives when compared to the other kernel functions. The radial basis kernel function is given in (6) as

$$\varnothing(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \text{ width parameter } \sigma > 0 \quad (6)$$

For the SVR path loss regression problem, we measure the difference between the actual and the predicted value by calculating the error of approximation because we are dealing with a regression and not classification problem.

We adopt the Vapnik-Chervonenkis principle to estimate the loss function with the E sensitivity zone expressed in (7) as

$$E(e_i) = |e_i|_\varepsilon = \begin{cases} 0, & \text{if } |e_i| \leq \varepsilon \\ |e_i| - \varepsilon, & \text{otherwise} \end{cases} \quad (7)$$

The error function e_i is the difference between the measured path loss and the predicted path loss from the SVR model and given in (8) as

$$e_i = y_i - h(x_i, w) \quad (8)$$

where y_i is the measured path loss and $h(x_i, w)$ is the predicted path loss from SVR model.

The last stage of the SVR model development is to accurately find a linear function that correctly maps the input to the output function. The function will approximate the training pairs (x_i, y_i) with the greatest accuracy.

The optimization function w that minimizes the error function is expressed in Equation (9) as

$$\text{Min}_{w,b} \frac{1}{2} \|w\|^2 \quad (9)$$

3.2. The Radial Basis Function Model Design

The radial basis function is essentially good for function approximation and

highly suitable for solving path loss regression problems. It is a three layer feed-forward network consisting of input layer, hidden layer and the output layer. The hidden layer performs a non-linear transformation of the input parameters while the output is a linear combiner of the outputs of the hidden layer. The radial basis function is efficient because the training is faster as it has one single hidden layer and so reduces all forms of complexities. The function of each node can also be easily interpreted in the RBF network.

For the RBF network, there are five input parameters into the network are elevation, clutter heights, distance between the transmitter and the receiving antenna, altitude, and building to building distance.

The objective of the radial basis function is to obtain a function that will give the path loss close to the measured path loss.

For data training and learning with the Gaussian kernel function. The output weighted path loss training points $\{t^p\}$. The radial basis function neural network $f(x)$ developed using the Gaussian kernel function as the weighted sum of Gaussian is given in (10).

$$f(x) = \sum_{n=1}^K W_p \phi_p(x) = \sum_{n=1}^K w_p \exp\left(-\left\|\frac{x-v^n}{2\sigma^2}\right\|\right) \quad (10)$$

σ is the width of the Gaussians, w represents the weight of the function, ϕ is the interpolation matrix, The distances $\|x-v^n\|$ give the Euclidean distance between input vector x and the center of the n^{th} neuron in the hidden layer.

Minimization in the RBF achieved as given in (11)

$$E = \frac{1}{N} \sum_{i=1}^N \|y_p - w^T \phi_p\|^2 \quad (11)$$

where $y = [y_1, y_2, \dots, y_N]$.

Hyperparameter Tuning of the RBF Model

The parameters of the radial basis function were tuned in order obtain optimal parameters for modelling. The Gaussian, multiquadric, inverse multiquadric functions were examined to determine which one is best for the model. The number of centroids for the particular kernel function is varied to determine the level of convergence. The number of centroids is plotted against the mean square error. The MSE value decreases as the number of centroids increases for all the kernel function that was explored. The Gaussian kernel function also gave the best performance. The hyperparameter tuning is presented in **Figure 3**.

Figure 3 strongly reflects that the value of error decreases as the number of centroids increases for a Gaussian kernel function.

The performances of five existing propagation models were compared with the introduced machine learning models using the experimental data. The models are Stanford University Interim (SUI) model, free space model, Cost-231 Walfisch Ikegami model, Cost-231 Hata model and Egli model. The model formulations are presented in Equations (12)-(16).

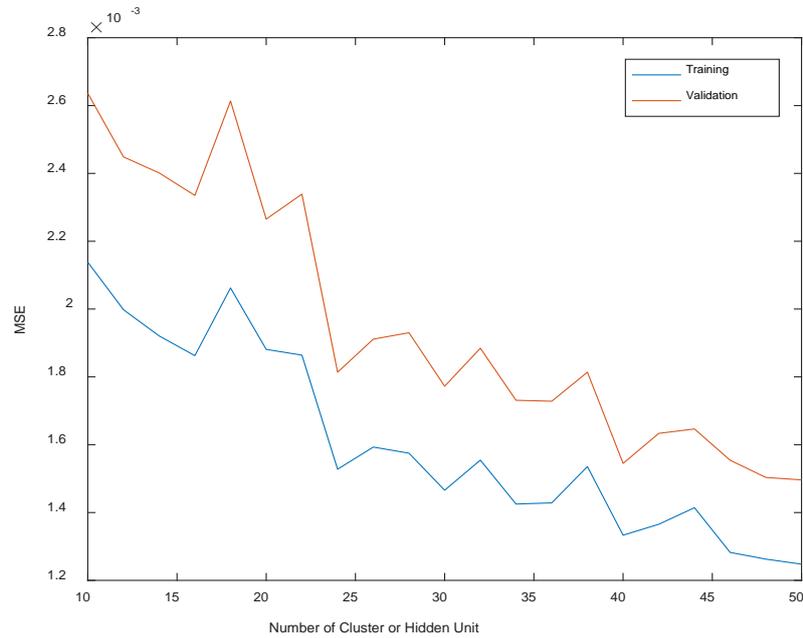


Figure 3. Plot of the MSE for the training and validation datasets for Gaussian Kernel function, Cluster range 10:2:50 and 10 runs.

$$\text{Free space } P_L \text{ (dB)} = 32.5 + 20 \log_{10}(d) + 20 \log_{10}(f) \tag{12}$$

$$\begin{aligned} \text{COST-231 Hata } P_L \text{ (dB)} &= 46.3 + 33.99 \log_{10}(f) - 13.82 \log_{10}(h_b) - ah_m \\ &+ [44.9 - 6.55 \log_{10}(h_b)] \log_{10} d + C_m \end{aligned} \tag{13}$$

$$ah_m = (1.1 \log_{10} f - 0.7) h_r - (1.56 \log_{10} f - 0.8) \tag{14}$$

$$\text{Eg}li = 40 \log_{10}(d) + 20 \log_{10}(f) - 20 \log_{10}(h_b) + L_m \tag{15}$$

$$\text{SUI} = A + \log_{10} \left(\frac{d}{d_0} \right) + K_f + K_h + S_f \text{ for } d \succ d \tag{16}$$

where d is the distance in km, f is the frequency in MHz, h_b represents the height of the base station antenna, h_r stands for the mobile antenna height, ah_m is the receiver antenna correction factor and parameter.

The developed machine learning models were validated using key performance metrics of root mean squared error (RMSE) and squared R. This is done to make a proper validation of the ML models and the existing analytical models.

$$\text{RMSE} = \sqrt{\sum_{i=1}^N \left(\frac{PL_M - PL_E}{N} \right)^2} \tag{17}$$

$$\text{Squared R} = 1 - \frac{\sum_{i=1}^N (PL_M - PL_E)^2}{\sum_{i=1}^N (PL_M - \overline{PL_M})^2} \tag{18}$$

PL_M represents measured path loss, PL_E represents the predicted path loss from the, $\overline{PL_M}$ is the mean of the measured path loss and N is the number of sampled data.

4. Results and Discussion

The performances of the two machine learning models, SVR and RBF were compared with the experimental data. The performances of SUI, Cost-231 Hata, Egli, freespace, Cost-231 Walfisch Ikegami models were also examined in comparison with the measured data. The results were examined in rural, suburban and urban areas.

Figure 4 gives the comparison of the measured path loss with the predicted path loss for the proposed SVR model in rural settings. It is clearly evident from the plots that the SVR model aligned well with the measured path loss. The SVR model fits the datasets accurately.

Figure 5 gives the plots of the comparison of the experimental data, SVR model

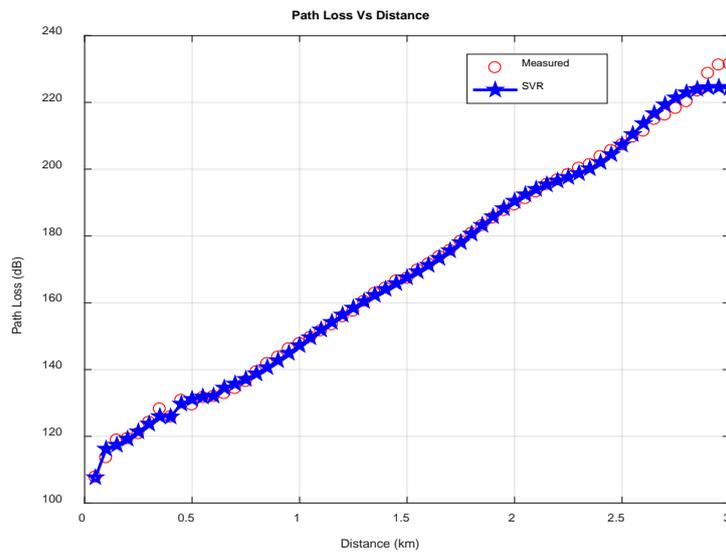


Figure 4. Plots of measured data and SVR model in rural settings.

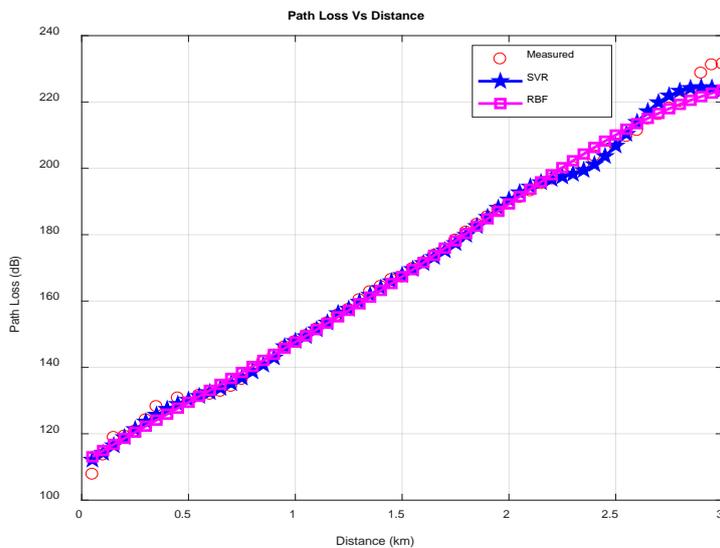


Figure 5. Plots of measured data, SVR and RBF models in rural settings.

and the RBF models. The plot also shows that the SVR model performed well than the RBF model. The SVR model aligned more accurately with the measured data than the RBF model.

Figure 6 presents the plots of the SVR and RBF models with the other five existing empirical models. The introduced machine learning path loss models predicted path loss more accurately than the empirical models. The five other models overpredicted path loss. The support vector regression model gave path loss closest to the experimental data. It was closely followed by the RBF model.

The performances of the seven models are presented in **Table 2** using RMSE and Squared R as an error metrics. A model with the lowest value of RMSE is the one with the greatest accuracy. The SVR gives the lowest value of RMSE with a value of 1.3678 dB in the rural areas. The value of squared R closest to 1 represents high accuracy. The SVR model gave a value of squared R = 0.9786 dB better than

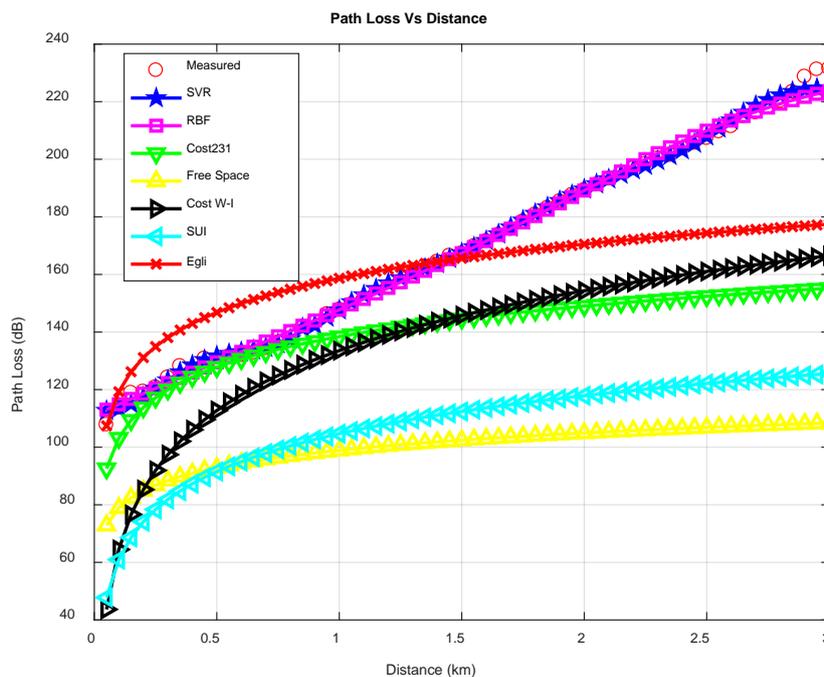


Figure 6. Plots of measured data, SVR, RBF and five empirical models in rural settings.

Table 2. Error metrics for different models in rural settings.

Models	RMSE value in (dB)	Squared R in (dB)
SVR	1.3678	0.9786
RBF	1.5765	0.9653
Cost-231 W-I	18.7652	0.3472
Egli	8.3592	0.7457
SUI	19.3894	0.3324
COST-231 Hata	16.6378	0.2441
Free Space	20.2341	0.1731

the other models and followed closely by the RBF model. From the RMSE and the squared R values, it shows that the empirical models predicted path loss outside of the given range and not suitable for signal propagation in the investigated environments.

4.1. Results for Areas Classified as Suburban

The results for the areas classified as suburban are presented in this section. The results obtained are similar to the results in the rural settings.

Figure 7 gives the comparison of the measured path loss with the predicted path loss for the proposed SVR model in suburban areas. It is clearly evident from the plots that the SVR model aligned well with the measured path loss. The SVR model fits the datasets accurately.

Figure 8 gives the plots of the comparison of the experimental data, SVR model and the RBF models in suburban areas. The plot also shows that the SVR model performed well than the RBF model. The SVR model aligned more accurately with the measured data than the RBF model. The results are also consistent with the one obtained in the rural areas.

Figure 9 presents the plots of the SVR and RBF models with other five existing empirical models in suburban areas. The introduced machine learning path loss models predicted path loss more accurately than the empirical models. The five other models overpredicted path loss. The support vector regression model gave path loss closest to the experimental data. It was closely followed by the RBF model.

The performances of the seven models are presented in **Table 3** using RMSE and squared R as the error metrics. A model with the lowest value of RMSE is the one with the greatest accuracy. The SVR gives the lowest value of RMSE with

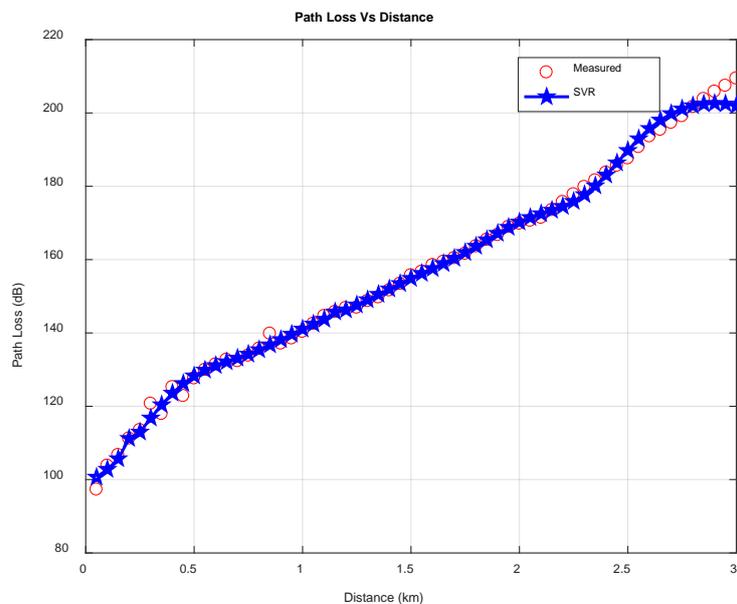


Figure 7. Plots of measured data and SVR model in suburban settings.

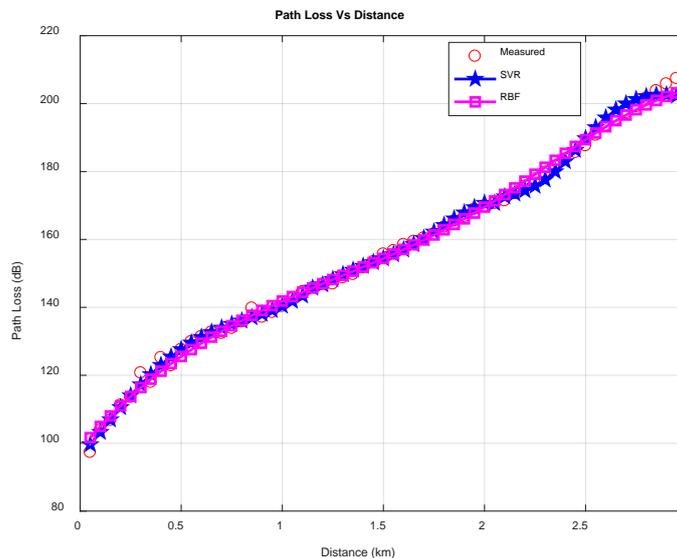


Figure 8. Plots of measured data, SVR and RBF models in suburban settings.

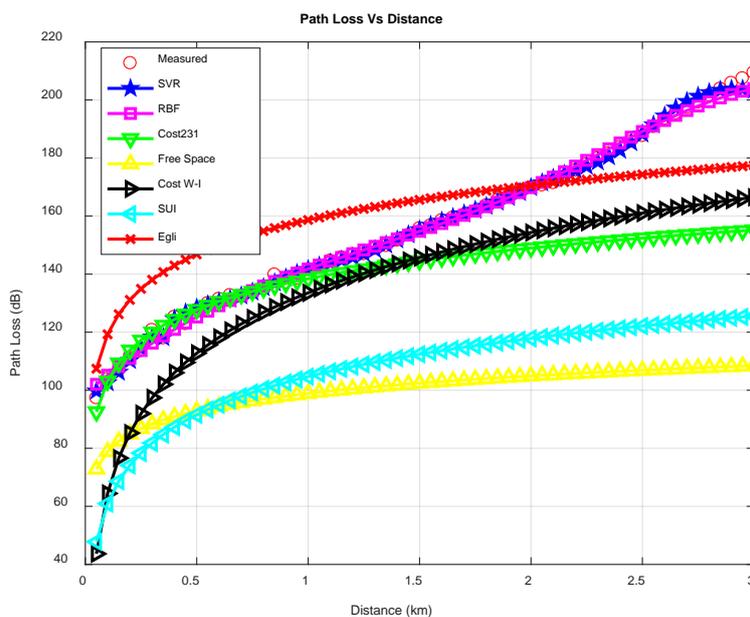


Figure 9. Plots of measured data, SVR, RBF and five empirical models in suburban settings.

Table 3. Error metrics values for different models in suburban settings.

Models	RMSE value in (dB)	Squared R in (dB)
SVR	1.4523	0.9286
RBF	1.6965	0.89653
Cost-231 W-I	19.4652	0.4471
Egli	9.3492	0.6457
SUI	21.3892	0.3324
COST-231 Hata	15.6374	0.2743
Free Space	22.2341	0.1273

a value of 1.4523 dB in the suburban areas. The value of squared R closest to one represents high accuracy. The SVR model gave a value of squared R = 0.9286 dB which is better than the other models and followed closely by the RBF model. From the RMSE and the squared R values of the plots, it showed that they predicted path loss outside of the given range. The results obtained for suburban areas were consistent with the ones obtained for rural areas. The results for rural areas were slightly better than the suburban areas because of the several NLOS situation obtained in these areas. There are so many LoS possibilities in the rural areas and so the path loss is well lower.

4.2. Results for Areas Classified as Urban

The results for the areas classified as urban are presented in this section. The results obtained are similar to the results in the rural and suburban areas.

Figure 10 gives the comparison of the measured path loss with the predicted path loss for the proposed SVR model in urban settings. It is clearly evident from the plots SVR model aligned well with the measured path loss. The SVR model fits the datasets accurately for this area.

Figure 11 gives the plots of the comparison of the experimental data, SVR model and the RBF models in urban areas. The plot also showed that the SVR model performed well than the RBF model. The SVR model aligned more accurately with the measured data than the RBF model. The results are also consistent with the one obtained in the rural and suburban areas.

Figure 12 presents the plots of the SVR and RBF models with other five existing empirical models in urban areas. The introduced machine learning path loss models predicted path loss more accurately than the empirical models. The five other models overpredicted path loss. The support vector regression

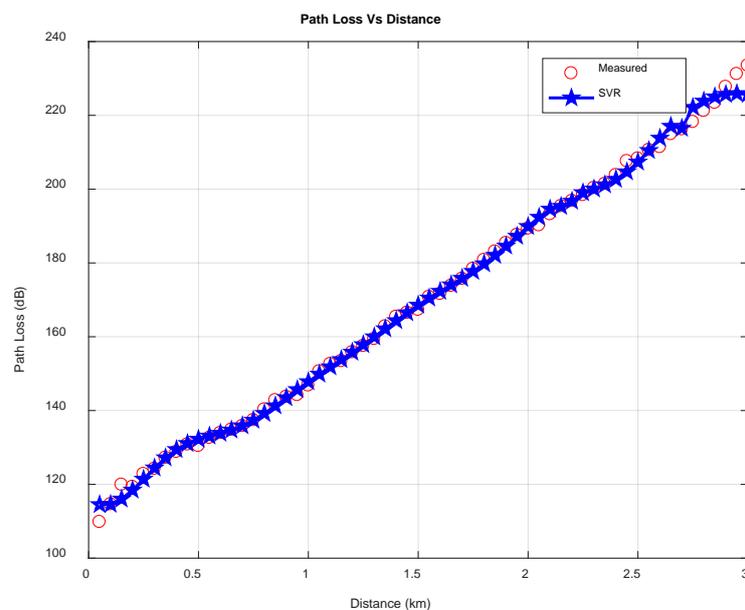


Figure 10. Plots of measured data and SVR model in urban settings.

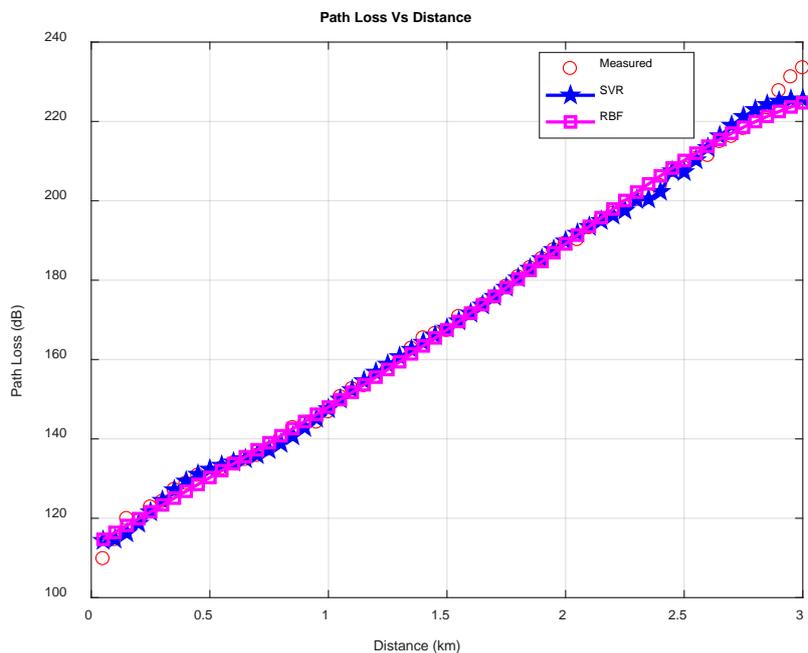


Figure 11. Plots of measured data, SVR and RBF models in urban settings.

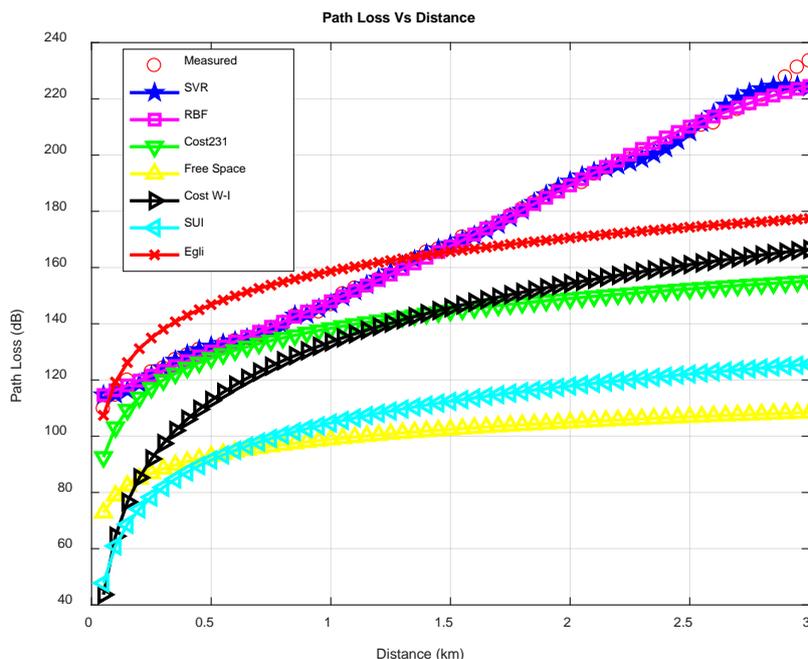


Figure 12. Plots of measured data, SVR, RBF and five empirical models in urban settings.

model gave path loss closest to the experimental data. It is closely followed by the RBF model. The path loss is slightly higher in the urban areas as shown because of the several multipath components. Signal is basically propagating under NLoS conditions in urban areas.

The performances of the seven models are presented in Table 4 using RMSE and squared R as error metrics. A model with the lowest value of RMSE is the

Table 4. Error metrics values for different models in urban settings.

Models	RMSE value in (dB)	Squared R in (dB)
SVR	2.1568	0.8528
RBF	2.4967	0.89653
Cost-231 W-I	22.4452	0.5671
Egli	8.3492	0.7457
SUI	21.3892	0.4324
COST-231 Hata	17.5384	0.5643
Free Space	22.2642	0.4753

one with the greatest accuracy. The SVR gives the lowest value of RMSE with a value of 2.1568 dB in the urban areas. The value of squared R closest to one represents high accuracy. The SVR model gave a value of squared R to be 0.8528 dB, which is better than the other models and followed closely by the RBF model. From the RMSE and the squared R, it showed that the other analytical models predicted path loss outside of the given range. The results obtained for urban areas were consistent with the ones obtained for rural and suburban areas. The results for rural areas were slightly better than the urban areas because of the several NLOS situation obtained in these areas. There are so many LoS possibilities in the rural areas and so the path loss is well lower.

5. Conclusion

The study introduced machine learning algorithms to path loss predictions with experimental data collected in Cyprus. Support vector regression (SVR) and radial basis function (RBF) are the two models that were trained with extensive field measurements. The predictive accuracies of five other empirical/analytical models were then analysed and compared with the performances of the machine learning models. The hyperparameters of the SVR and the RBF models were tuned before modeling and the Gaussian kernel function gave the best performing results; it was subsequently adopted and used in the ML models. The validation was then carried out with the use of root-mean square error (RMSE) and squared R (R^2). The SVR and the RBF models gave the lower values of RMSE than the other five empirical models. The SVR model performed best and has values of 1.3678 dB, 1.4523 dB, 2.1568 dB in rural, suburban and urban settings respectively. The other empirical models namely SUI, Egli, Cost-231, Cost-231 W-I, Freespace overpredicted path loss and are therefore not suitable for signal propagation in these environments. The introduced ML algorithms (SVR and RBF) aligned accurately with the measured data. They gave a high predictive accuracy. This is the power of machine learning algorithms over analytical models. The SVR gave the highest predictive accuracy because it trains the model using a symmetrical loss function and its computational complexity does not depend on the dimensions of input parameters. Conclusively, we have been able to demon-

strate that machine learning models are more efficient for path loss prediction than the analytical models. Consequently, signal propagation, interference analysis, network planning and cell parameters evaluation should be carried out with these Machine learning models (SVR and RBF) because they produced higher predictive accuracies as demonstrated in this study.

Data Availability Statement

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] Xing, Y. and Rappaport, T.S. (2018) Propagation Measurement System and Approach at 140 GHz-Moving to 6G and above 100 GHz. *2018 IEEE Global Communications Conference (GLOBECOM)*, Dhabi, 9-13 December 2018, 1-6. <https://doi.org/10.1109/GLOCOM.2018.8647921>
- [2] Wu, J., Guo, S., Huang, H., Liu, W. and Xiang, Y. (2018) Information and Communications Technologies for Sustainable Development Goals: State-of-the-Art, Needs and Perspectives. *IEEE Communications Surveys & Tutorials*, **20**, 2389-2406. <https://doi.org/10.1109/COMST.2018.2812301>
- [3] Oseni, O.F., Popoola, S.I., Enumah, H. and 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**, 174-180. <https://doi.org/10.15623/ijret.2014.0308027>
- [4] Obot, O.S. and Afolayan, J. (2011) Comparative Analysis of Path Loss Prediction Models for Urban Macrocellular Environments. *Nigerian Journal of Technology*, **30**, 50-59.
- [5] Popoola, S.I., Atayero, A.A. and Popoola, O.A. (2018) Comparative Assessment of Data Obtained Using Empirical Models for Path Loss Predictions in a University Campus Environment. *Data Brief*, **18**, 380-393. <https://doi.org/10.1016/j.dib.2018.03.040>
- [6] Faruk, N., Popoola, S.I., Surajudeen-Bakinde, N.T., Oloyede, A.A., Abdulkarim, A., Ali, M., Calafate, C.T., Atayero, A.A. and Olawoyin, L.A. (2019) Path Loss Predictions in the VHF and UHF Bands within Urban Environments: Experimental Investigation of Empirical, Heuristics and Geospatial Models. *IEEE Access*, **7**, 77293-77307. <https://doi.org/10.1109/ACCESS.2019.2921411>
- [7] Popoola, S.I., Adetiba, E., Atayero, A.A., Faruk, N., Yuen, C. and Calafate, C.T. (2018) Optimal Model for Path Loss Predictions Using Feed-Forward Neural Networks. *Cogent Engineering*, **5**, Article No. 1444345. <https://doi.org/10.1080/23311916.2018.1444345>
- [8] Salman, M.A., Popoola, S.I., Faruk, N., Surajudeen-Bakinde, N.T., Oloyede, A.A., and Olawoyin, L.A. (2017) Adaptive Neuro-Fuzzy Model for Path Loss Prediction in the VHF Band. *2017 International Conference on Computing Networking and Informatics (ICCNi)*, Lagos, 29-31 October 2017, 1-6. <https://doi.org/10.1109/ICCNi.2017.8123768>

- [9] Noh, S.-K., Kim, G.-S., Kim, S.W. and Choi, D. (2014) Measurements and Analysis for the Proposal of Propagation Models Considering the Time and Spatial Variation of Propagation Environments in the Micro-Wave Band. 2014 *International Conference on Information and Communication Technology Convergence (ICTC)*, Busan, 22-24 October 2014, 456-461. <https://doi.org/10.1109/ICTC.2014.6983180>
- [10] Har, D., Xu, C. and Xia, H.H. (2000) Propagation Modelling for Wireless Local Loop Channel. *International Journal of Communication Systems*, **13**, 231-241. [https://doi.org/10.1002/1099-1131\(200005\)13:3<231::AID-DAC428>3.0.CO;2-Q](https://doi.org/10.1002/1099-1131(200005)13:3<231::AID-DAC428>3.0.CO;2-Q)
- [11] Sun, R. and Matolak, D.W. (2017) Path Loss and Delay Spread for the Stairwell channel at 5GHz. *International Journal of Communication Systems*, **30**, Article No. e2920. <https://doi.org/10.1002/dac.2920>
- [12] Hong, T., Ku, B., Park, J., Ahn, D.-S. and Jang, Y.-S. (2006) Capacity of the WCDMA System Using High Altitude Platform Stations. *International Journal of Wireless Information Network*, **13**, 5-17. <https://doi.org/10.1007/s10776-005-0014-x>
- [13] Park, J.-J., Kim, M.-D., Kwon, H.-K., Chung, H.K., Yin, X. and Fu, Y. (2012) Measurement-Based Stochastic Cross-Correlation Models of a Multilink Channel in Cooperative Communication Environments. *ETRI Journal*, **34**, 858-868. <https://doi.org/10.4218/etrij.12.1812.0039>
- [14] Faruk, N., Ayeni, A. and 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>
- [15] Dai, H.-N., Na, K.-W., Li, M. and Wu, M.-Y. (2013) An Overview of Using Directional Antennas in Wireless Networks. *International Journal of Communication Systems*, **26**, 413-417. <https://doi.org/10.1002/dac.1348>
- [16] Zhang, Y., Wen, J., Yang, G., He, Z. and Wang, J. (2019) Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion. *Applied Sciences*, **9**, Article No. 1908. <https://doi.org/10.3390/app9091908>
- [17] Piacentini, M. and Rinaldi, F. (2011) Path Loss Prediction in an Urban Environment Using Learning Machines and Dimensionality Reduction Techniques. *Computational Management Science*, **8**, Article No. 371. <https://doi.org/10.1007/s10287-010-0121-8>
- [18] Ubom, E.A., Idigo, V., Azubogu, A.C., Ohaneme, C. and Alumona, T.L. (2011) Path Loss Characterization of Wireless Propagation for South-South Region of Nigeria. *International Journal of Computer Theory and Engineering*, **3**, 360-364. <https://doi.org/10.7763/IJCTE.2011.V3.332>
- [19] Erceg, V., Fortune, S.J., Ling, J., Rustako, A.J. and Valenzuela, R.A. (1997) Comparisons of a Computer-Based Propagation Prediction Tool with Experimental Data Collected in Urban Microcellular Environments. *IEEE Journal on Selected Areas in Communications*, **15**, 677-684. <https://doi.org/10.1109/49.585778>
- [20] Neskovic, A., Neskovic, N. and Paunovic, G. (2000) Modern Approaches in Modeling of Mobile Radio Systems Propagation Environment. *IEEE Communications Surveys and Tutorials*, **3**, 2-12. <https://doi.org/10.1109/COMST.2000.5340727>
- [21] Bakinde, N.T.S., Faruk, N., Ayeni, A.A., Muhammad, M.Y. and Gumel, M.I. (2012) Comparison of Propagation Models for GSM 1800 and WCDMA Systems in Selected Urban Areas of Nigeria. *International Journal of Applied Information Systems*, **2**, 6-13.
- [22] Emagbetere, J.O. and Edeko, F.O. (2009) Measurement Validation of Hata-Like Models for Radio Propagation Path Loss in Rural Environments at 1.8GHz. *Journal of Mobile Communications*, **3**, 17-21.

- [23] Kalakh, M., Kandil, N. and Hakem, N. (2012) Neural Networks Model of a UWB Channel Path Loss in a Mine Environment. *2012 IEEE 75th Vehicular Technology Conference (VTC Spring)*, Yokohama, 6-9 May 2012, 1-5. <https://doi.org/10.1109/VETECS.2012.6240318>
- [24] Cerri, G., Cinalli, M., Michetti, F. and Russo, P. (2004) Feedforward Neural Networks for Path Loss Prediction in Urban Environment. *IEEE Transactions on Antennas and Propagation*, **52**, 3137-3139. <https://doi.org/10.1109/TAP.2004.835252>
- [25] Sotiroudis, S.P., Goudos, S.K., Gotsis, K.A., Siakavara, K. and Sahalos, J.N. (2013) Application of a Composite Differential Evolution Algorithm in Optimal Neural Network Design for Propagation Path-Loss Prediction in Mobile Communication Systems. *IEEE Antennas and Wireless Propagation Letters*, **12**, 364-367. <https://doi.org/10.1109/LAWP.2013.2251994>
- [26] Han, S. and Kim, H. (2019) On the Optimal Size of Candidate Feature Set in Random Forest. *Applied Sciences*, **9**, Article No. 898. <https://doi.org/10.3390/app9050898>
- [27] Imoize, A. (2013) Propagation Measurements and Modelling 1800 MHz in Lagos, Nigeria. *International Journal of Wireless Communications and Mobile Computing*, **6**, 165-173. <https://doi.org/10.1504/IJWMC.2013.054042>
- [28] Ojo, S., Imoize, A. and Alienyi, D. (2020) Radial Basis Function Neural Network Path Loss Prediction Model for LTE Networks in Multitransmitter Signal Propagation Environments. *International Journal of Communication Systems*, **34**, Article No. e4680. <https://doi.org/10.1002/dac.4680>
- [29] Ojo, S., Akkaya, M. and Suporu, J. (2022) An Ensemble Machine Learning Approach for Enhanced Path Loss Predictions. *International Journal of Communication Systems*, **35**, Article No. e5101. <https://doi.org/10.1002/dac.5101>
- [30] Bilgehan, B. and Ojo, S. (2018) Multiplicative Based Path Loss Model. *International Journal of Communication Systems*, **31**, Article No. e3794. <https://doi.org/10.1002/dac.3794>
- [31] Isabona, J., Imoize, A.L., Ojo, S., Lee, C.-C. and Li, C.-T. (2022) Atmospheric Propagation Modelling for Terrestrial Radio Frequency Communication Links in a Tropical Wet and Dry Savanna Climate. *Information*, **13**, Article No. 141. <https://doi.org/10.3390/info13030141>
- [32] Isabona, J., Imoize, A.L., Rawat, P., Jamal, S.S., Pant, B., Ojo, S. and Hinga, S.K. (2022) Realistic Prognostic Modeling of Specific Attenuation due to Rain at Microwave Frequency for Tropical Climate Region. *Wireless Communications and Mobile Computing*, **2022**, Article ID: 8209256, 10 p. <https://doi.org/10.1155/2022/8209256>
- [33] Olawumi, T.O., Ojo, S., Chan, D.W.M. and Yam, M.C.H. (2021) Factors Influencing the Adoption of Blockchain Technology in the Construction Industry—A System Dynamics Approach. *CRIOCM 2020: Proceedings of the 25th International Symposium on Advancement of Construction Management and Real Estate*, Wuhan, 28-30 November 2020, 1235-1249. https://doi.org/10.1007/978-981-16-3587-8_84
- [34] Olawumi, T.O., Chan, D.W.M., Ojo, S. and Yam, M.C.H. (2022) Automating the Modular Construction Process: A Review of Digital Technologies and Future Directions with Blockchain Technology. *The Journal of Building Engineering*, **46**, Article ID: 103720. <https://doi.org/10.1016/j.jobbe.2021.103720>
- [35] Kavitha, K.V.N., Ashok, S., Imoize, A.L., Ojo, S., Senthamil Selvan, K., Ahanger, T.A. and Alhassan, M. (2022) On the Use of Wavelet Domain and Machine Learning for the Analysis of Epileptic Seizure Detection from EEG Signals. *Journal of*

Healthcare Engineering, **2022**, Article ID: 8928021, 16 p.

<https://doi.org/10.1155/2022/8928021>

- [36] Isabona, J., Imoize, A.L., Ojo, S., Karunwi, O., Kim, Y., Lee, C.-C. and Li, C.-T. (2022) Development of a Multilayer Perceptron Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments. *Applied Sciences*, **12**, Article No. 5713. <https://doi.org/10.3390/app12115713>

List of Abbreviations

SVR	Support vector regression
RBF	Radial basis function
ML	Machine learning
BTS	Base transceiver station
LoS	Line of sight
Nlos	Non-line of sight
TEMs	Test mobile system
RMSE	Root mean squared error
MSE	Mean square error