Wiley
International Journal of Antennas and Propagation
Volume 2025, Article ID 3277479, 17 pages
https://doi.org/10.1155/ijap/3277479

Research Article

A Novel Hybrid Path Loss Prediction Model for 5G Midband Networks Using Empirical, Machine Learning, and Feature Prioritization Techniques

Farouq E. Shaibu,[1] **Elizabeth N. Onwuka**,[1] **Nathaniel Salawu**,[1] **Stephen S. Oyewobi**,[1] **Adnan M. Abu-Mahfouz**,[2,3] **and Karim Djouani**[2,4]

[1]*Department of Telecommunication Engineering, Federal University of Technology Minna, Minna 920211, Nigeria*
[2]*Tshwane University of Technology, Pretoria 0001, South Africa*
[3]*Council for Scientific and Industrial Research, Pretoria 0083, South Africa*
[4]*LISSI Laboratory, Universite Paris-Est Creteil Val de Marne Bibliotheques, Créteil 94000, France*

Correspondence should be addressed to Adnan M. Abu-Mahfouz; aabumahfouz@csir.co.za

Received 7 December 2024; Revised 3 September 2025; Accepted 26 September 2025

Academic Editor: Venkata Kishore Kothapudi

Accurate path loss prediction is vital for 5G deployment, especially at midband frequencies where signal degradation is significant. This paper presents a hybrid model that integrates an optimized COST-231 Hata model with a random forest algorithm to improve prediction accuracy at 3.5 GHz. Recursive feature elimination identified eleven key features from eighteen multidimensional parameters, including novel environmental attributes, to prioritize factors influencing urban path loss. Validation against measurement and simulation datasets showed strong alignment with observed results, achieving lower errors (MAE = 1.82 dB, RMSE = 2.05 dB, and MAPE = 2.4%) compared to existing models. Additionally, cross-band validation at 1.6 GHz further demonstrated the model's robustness, though retraining or fine-tuning is recommended for optimal performance at lower frequencies. Future research may expand the dataset to enhance generalizability.

Keywords: 5G communication; COST-231 Hata model; feature selection; hybrid model; path loss; random forest model

1. Introduction

The advent of fifth-generation (5G) wireless communication networks has introduced revolutionary changes in various sectors, addressing critical issues such as data speed, latency, and connectivity. However, despite these improvements, 5G networks come with their own set of challenges, particularly in midband frequencies like 3.5 GHz, where path loss remains a significant concern. Path loss is not particular to 3.5 GHz or 5G midband, because the higher the frequency, the more the path loss challenge. However, the midband is on the front burner because of its strategic position on the spectrum. It is between the long wavelength and the short waves; therefore, it can fit in for backhauling and for access networks. Therefore, working in the midband is beneficial either way, making it a priority band. Path loss affects other aspects of the network, such as power consumption, coverage, interference, infrastructure costs, increased latency, and a higher bit error rate (BER). Addressing path loss is crucial, as it not only enhances network reliability but also optimizes overall performance in wireless communication systems, making it a priority in the development of robust 5G networks.

Several empirical models have been developed over time to predict path loss under different conditions, such as the widely used COST-231 Hata model, originally designed for frequencies below 2 GHz. However, with 5G operating in higher-frequency ranges, existing models often fail to deliver the required accuracy. Machine learning techniques have recently shown potential in improving path loss prediction accuracy by learning complex patterns from large datasets. However, a major gap remains in how to efficiently combine the strength of empirical models and machine learning techniques to deliver a more accurate and adaptable solution, especially in midband frequencies [1, 2].

This paper addresses the limitations of traditional empirical models and standalone machine learning models by developing a novel hybrid approach. It starts by optimizing the COST-231 Hata model to align with the characteristics of 5G midband at 3.5 GHz. Subsequently, a random forest (RF) model is trained using the best eleven features selected by the recursive feature selection technique. These features include factors influencing signal propagation, which are then prioritized for their significance in determining path loss [2]. The trained RF model is employed to refine the optimized COST-231 Hata model, resulting in a hybrid model that leverages both empirical insights and the predictive power of machine learning. This hybrid model aims to improve prediction accuracy, reduce path loss, and address the key challenges of 5G networks.

Many countries across the globe have designated the 3.5-GHz frequency band for implementing 5G networks. Aligning with this priority of midband frequency spectrum for the adoption of 5G communication infrastructure, the Nigerian Communication Commission (NCC) has allocated the midband spectrum ranging from 3.4 to 3.8 GHz for the deployment of 5G communication technology in Nigeria. Specific usage scenarios for this spectrum allocation are delineated in Figure 1.

The main contributions of this paper include the following:

i. Optimization of the COST-231 Hata model for 5G midband frequency at 3.5 GHz.
ii. Integration of machine learning techniques, specifically the RF algorithm, to enhance path loss prediction.
iii. Feature selection and prioritization to identify the most significant factors influencing path loss in 5G networks.
iv. Development of a novel hybrid model that synergizes empirical and machine learning approaches for improved accuracy and performance.
v. Evaluation of the model's performance with an independent dataset from Ogun State, Nigeria, at 1.6 GHz, demonstrating its effectiveness and generalization.

The remainder of this paper is structured as follows: Section 2 discusses empirical and machine learning path loss models. Section 3 reviews related work on path loss modeling and machine learning approaches in wireless networks. Section 4 outlines the methodology used in optimizing the COST-231 Hata model and developing the RF-based hybrid model. Results and discussion, as well as the validation of the models, are presented in Section 5. Finally, Section 6 discusses the implications of the findings and concludes with potential areas for future research.

2. Path Loss Model

The free space path loss (FSPL) model is often used as a baseline reference for estimating the path loss in wireless communication systems, especially for line-of-sight (LoS) scenarios.

Typically, the attenuation of propagation paths is influenced by both the frequency of the signal and the distance it travels, expressed mathematically as follows:

$$P_l = 10\log_{10}\left(\left(\frac{4\pi d f}{c}\right)^2\right), \tag{1}$$

where P_l, f, c, and d represent the path loss, frequency of operation, the velocity of propagation in free space, and the distance between the transmitter and the receiver, respectively.

As we can see, (1) can thus be transformed as follows [3]:

$$P_l = 92.45 + 20\log_{10}(d) + 20\log_{10}(f), \tag{2}$$

where f and d represent the frequency in megahertz and the path length between the Tx and Rx in kilometers.

To compare the measured path loss to the predicted path loss, the measured path loss can thus be presented as follows [4]:

$$P_l = \mathrm{EIRP} - R_p, \tag{3}$$

where R_p and EIRP are the received power and the effective isotropic radiated power, which are determined by several factors including the transmitting power, P_T (dBm), the gains of the transmitting antenna G_T and receiving antenna G_R, and potential losses attributable to the feeder cable, F_l, antenna, A_l, and filter, A_{fl}, illustrated as follows:

$$\mathrm{EIRP} = P_T + G_T + G_R - F_l - A_l - A_{fl}. \tag{4}$$

However, in real-world environments, the path loss may be affected by various factors such as terrain, obstacles, reflections, and diffractions, which can cause additional attenuation and affect the overall signal strength, as illustrated in Figure 2 [5].

2.1. Empirical Path Loss Model.

Empirical channel models for path loss are statistical models that capture the attenuation of radio signals as they propagate through various environments. These models are derived from extensive measurements of real-world wireless communication scenarios and are essential for predicting signal strength and optimizing network design. Many empirical models have been developed in the measured environment, but they fail when applied to different measurement environments and experimental setups for which they were developed [6, 7].

Several propagation models based on empirical measurements have been proposed in the literature, including the Okumura model [8], the Okumura–Hata model [9], the COST-231 Hata model [10], and the Walfisch–Ikegami model [11]. Among these options, the COST-231 Hata model is selected for this study due to its proven accuracy

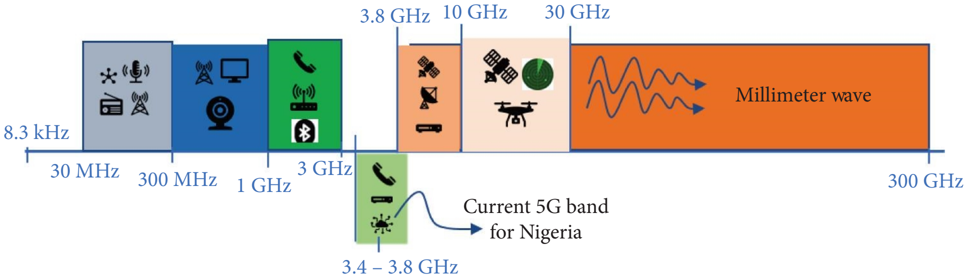


Figure 1: Nigeria's frequency spectrum.

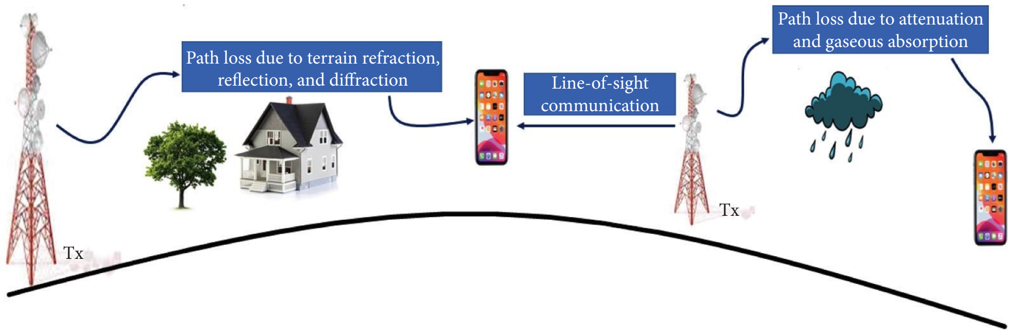


Figure 2: Propagation mechanism effect along a terrestrial path.

and reliability, especially in the frequency range of interest (2 GHz) [12]. This model, an extension of the Okumura–Hata model [11], is preferred for its robustness and use of extensive and diverse measurement data. Moreover, it enjoys wide acceptance and recognition by governmental agencies and international standardization bodies.

The model provides a formula to calculate path loss between a transmitter and receiver, considering variables such as signal frequency, distance between the two, and environmental factors [13]. The model includes separate components for both LoS and non-LoS (NLoS) transmission and can be adapted for use in different frequency ranges and environmental conditions [9]. The conventional COST-231 Hata model [2] is depicted in (5), alongside its correction factor illustrated as follows:

$$PL = 46.33 + 33.9\log f - 13.82\log h_{tx} + C_m - a(h_{rx}) + (44.99 - 6.55\log(h_{tx}))\log d, \tag{5}$$

where PL, f, h_{tx}, h_{rx}, and d are the path loss in dB, frequency in MHz, transmitter antenna height in meters, receiver antenna height in meters, and the distance from the transmitting station to the receiving antenna in km. The correction factor is thus written as follows [14]:

$$a(h_{rx}) = (1.1\log_{10}(f) - 0.7)h_{rx} - (1.56\log_{10}(f) - 0.8), \quad \text{for } 1500\,\text{MHz} \leq f \leq 2000\,\text{MHz} \tag{6}$$

The COST-231 Hata model holds significant usage within the telecommunications sector for path loss estimations and the planning of mobile communication infrastructures within urban locales. Nevertheless, it is crucial to acknowledge that the model possesses certain constraints and may not invariably provide precise path loss predictions across all particular circumstances [14]. Hence, there is a need to optimize the model to suit a scenario with a frequency of interest of 3.5 GHz, which is 3.5 GHz.

2.2. Machine Learning Models. Machine learning path loss models are predictive models used in communication networks to estimate the path loss between wireless devices [15]. It leverages machine learning algorithms to analyze and learn from historical data, enabling the prediction of path loss in different environments [8]. These models undergo training using diverse parameters, including distance, topography, impediments, frequency, and antenna specifications, to predict the signal attenuation during its transmission from the transmitter to the receiver [16]. Machine learning models are classified into supervised, unsupervised, semi-supervised, and reinforcement machine learning [17].

Among some of the supervised machine learning models, such as linear regression, RF [18], K-nearest neighbors (k-NN) [19], extreme gradient boosting (XGBoost), and artificial neural network (ANN) [20], the RF model is considered for this paper. This is because the RF model reduces the risk of overfitting and the required training time. Additionally, it offers a high level of accuracy, as reported in lots of reviewed papers [21–26].

3. Related Works

The growth of wireless communication networks has led to an increased demand for accurate path loss models, particularly in urban environments where signal propagation is challenging. This section provides an overview of the existing literature on path loss modeling, focusing on prior works that have used machine learning techniques and the application of the COST-231 Hata model in urban environments at 3.5 GHz.

For example, the work in [27] proposed a deep learning–based path loss model that combines LoS and NLoS propagation scenarios while considering eight features for the modeling. The proposed model outperformed the two chosen current empirical models, such as the CI and ABG, that were compared, having a lower mean error. However, the stability of the proposed model is required to be tested in a more commercial environment. Further refinement is therefore required to improve the performance of the proposed model.

The authors of [28] used a model-aided deep learning approach to predict path loss at 7 GHz in an urban environment. Unlike previous models, their method integrates a modified set of input features and directly forecasts path loss instead of received signal strength. They leverage both modeling-based and learning-based techniques to enhance prediction accuracy, with numerical results indicating superiority over most empirical models. However, the proposed model requires further refinement, especially in environments with higher obstruction levels.

As reported in [29], machine learning methods, including the RF and NN models, exhibit the potential for enhancing the accuracy of path loss prediction within the midband frequency range of 1–6 GHz. Yet, research is scarce concerning the efficient integration of these methods into empirical models for path loss specifically within the 3.5-GHz band in urban environments.

In a comparable study conducted by [26], a promising approach to enhance path loss forecasting involved employing an RF model. Their experimentation revealed that integrating the RF method with attributes such as geographical coordinates, distance, azimuth, and antenna gain yielded superior outcomes compared to other models under consideration.

The study by [24] employed three machine learning models to analyze radio channel modeling in urban vehicular environments, focusing on LoS and NLoS scenarios using ray tracing–generated data. The RF model outperformed with a minimum root-mean-square error (RMSE) value of 1.80 ns, whereas the MLP model had a maximum RMSE of 11 ns. Conversely, the authors of [30] presented alternative methods to enhance path loss models by using machine learning techniques. They successfully reduced the number of measurements needed for modeling wireless channels through regression techniques.

In [25], the RF model underwent hyperparameter tuning to optimize path loss prediction in the Palembang City area of Indonesia. The model considered seven input features from a drive test measurement campaign, including cell distance, frequency, total height of Tx and Rx, vertical and horizontal angles of Rx from main beam Tx, road width, and adjacent building elevation. Results indicated improved accuracy, with the RF model achieving an 8.00% mean absolute percent error (MAPE) and an 11.8 dB RMSE, outperforming previous outcomes.

In a separate study conducted by [31], wideband measurements conducted within a street environment were used to establish path loss models for frequency bands spanning 1.8, 3.5, and 28 GHz. The measurement campaign took place under static environmental conditions without wind. The findings highlight the importance of considering multiple scattering effects from trees in the 1.8-GHz and 3.5-GHz bands after a certain distance between the transmitter and the receiver.

In their study, the authors of [32] enhanced the traditional COST-231 Hata model for the city of Limbe in Cameroon by employing optimization techniques such as the magnetic optimization algorithm (MOA) to predict path losses within the 2-GHz midband. The dataset used was obtained from drive tests conducted in the field. Although the proposed model yielded satisfactory prediction outcomes, it is noted that the MOA used may encounter challenges in addressing nonlinearities and could face convergence issues, potentially compromising the accuracy of the model.

In another study by [33], five distinct empirical models—FSPL [33], SUI model [33], Ericsson model, Okumura model [8], and COST-231 Hata model [10]—were evaluated against empirical data measurements to determine the most suitable model for predicting path loss in the urban setting of Cologne, Germany, operating at 2.5 GHz. The findings revealed that the COST-231 Hata model exhibited the best fit, achieving a minimum RMSE of 5.27 dB. However, it was noted that further refinement of the COST-231 Hata model is necessary to accommodate the specific characteristics of the urban environment, which includes a mix of old and modern buildings.

However, further fine-tuning based on the feature selection technique is required to improve the performance of the models in [24, 26, 28], especially in a cluttered urban environment. Also, further attributes should be incorporated to refine the models' performances through comprehensive training and testing. Extensive measurement campaigns across similar environments will be necessary in [31–33] to evaluate the stability of the models.

Therefore, in this paper, a supervised machine learning model (RF model) underwent refinement through the recursive feature elimination (RFE) technique. Subsequently, it was integrated with the fine-tuned and optimized COST-231 Hata model to develop a novel hybrid path loss model.

4. Methodology

4.1. Study Area. The study areas for the measurement campaigns, aimed at capturing received signal strength from two different mobile stations, are both located in the urban environment of Maitama, a district in the Federal Capital Territory of Nigeria [34]. These areas, identified by cellular site IDs FC0125 and FC3112, exhibit distinctive urban characteristics that contribute to the variation in signal propagation and reception.

The site designated as FC0125 is situated in a part of Maitama characterized by a well-planned and modern urban landscape. This area features a well-arranged configuration of tall, contemporary buildings that dominate the skyline. The architecture here primarily consists of high-rise residential and commercial structures constructed with modern materials and designed with a focus on vertical expansion. The streets in the vicinity are systematically laid out, offering a grid-like pattern that facilitates efficient navigation and urban mobility. These streets are wide, well-maintained, and equipped with modern amenities, including traffic management systems, pedestrian walkways, and landscaping elements such as trees and green spaces [34], as presented in Figure 3(a). The presence of such tall buildings and orderly streets creates a complex urban environment for signal propagation, with the potential for significant multipath reflections and signal diffraction.

The second site, FC3112, is located in a more diverse part of Maitama, where the urban landscape is a blend of both modern and older buildings, as shown in Figure 3(b). The modern buildings here are similar to those around FC0125, comprising tall structures with contemporary designs. However, mixed among these are older buildings, which add historical and architectural diversity to the area. The streets around FC3112 are less uniformly arranged compared to those around FC0125. This combination of modern and old buildings, along with the irregular street layout, creates a dynamic and heterogeneous environment for signal measurements, potentially influencing signal strength and propagation in different ways compared to the more uniformly modern area of FC0125.

Table 1 provides comprehensive technical and geographic parameters of the transmitting station used in this study. The table includes detailed technical parameters of the two selected base stations, along with descriptions of their respective locations.

4.2. Data Collection. The empirical data used in this study to develop a novel path loss model at 3.5 GHz in an urban environment were gathered through four different sources, which are presented as follows.

4.2.1. Measurement Campaigns. To conduct an outdoor measurement campaign at the 5G midband frequency of 3.5 GHz, extensive field measurement campaigns across the two diverse urban environments (regular and irregular urban) hosting operational 5G networks at the midband frequency of 3.5 GHz were conducted. In both regular and irregular urban environments, signal measurements were systematically repeated three times at each designated location, and the averages were calculated to guarantee precision and reliability in the collected data. A substantial number of measurements were undertaken across numerous points to ensure a comprehensive dataset for thorough analysis.

Measurements from the base station were conducted in a regular urban environment, identified as ID Number FC0125. Data were collected along two distinct paths: LoS and NLoS. The designated paths are visually represented in Figure 3(a), where the LoS path is marked with a circular shape filled in white and outlined in red, and the NLoS path is marked with a circular shape filled in black and outlined in yellow [34].

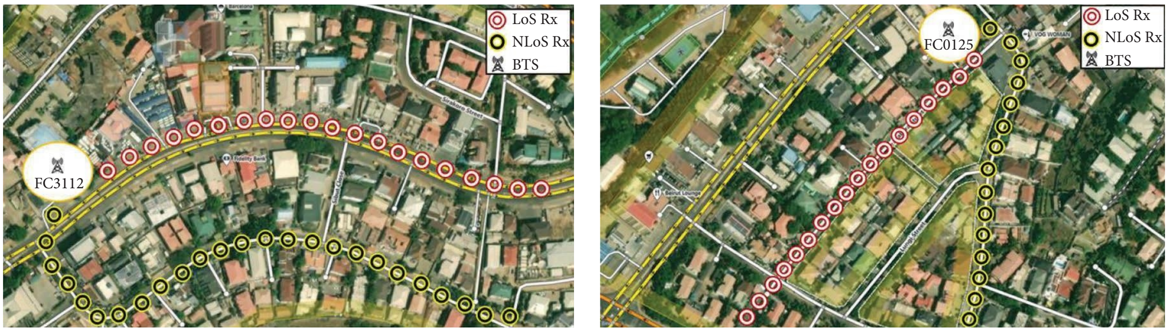
For reference, the distance between the tower leg and the palisade was determined to be 6 m. Using a tape rule, an additional 4 m was measured from the palisade to establish a reference point of 10 m from the tower leg. This reference point served as the starting distance for both LoS and NLoS paths. Subsequent measurements were taken at regular intervals of 10 m, extending up to a maximum distance of 710 m, aligning with the transmitter's coverage limit of 0.71 km. This systematic approach resulted in a total of 71 data points for each path. The measurement setup is shown in Figure 4.

Similarly, measurements were conducted from the base station in an irregular urban environment, identified as ID Number FC3112. Data collection was performed along two distinct paths: LoS and NLoS. These paths are visually represented in Figure 3(b), where the LoS path is marked with a circular shape filled in white and outlined in red, whereas the NLoS path is marked with a circular shape filled in black and outlined in yellow.

The distance between the tower leg and the palisade was measured to be 3 m. Using a tape rule, an additional 7 m was measured from the palisade to establish a reference point, creating a total distance of 10 m from the tower leg. This reference point was used for both the LoS and NLoS paths. Subsequent measurements were taken at regular intervals of 10 m, extending up to a maximum distance of 690 m, which corresponds to the transmitter's coverage limit of 0.69 km. This systematic approach resulted in a total of 69 data points for each path.

Figure 5 illustrates how these campaigns were carried out using a stationary base transmitter station, Tx, and a handheld spectrum analyzer (N9344C) as a directional receiver, Rx, which was mounted on a vehicle roof to the directional antenna (HE200) and driven along streets.

4.2.2. 3D Ray Tracing. WinProp (Altair HyperWorks) software tool was used to conduct 3D ray tracing. The study commenced by defining the geographical boundaries of the



(a) (b)

Figure 3: Study area: (a) regular urban environment and (b) irregular urban environment.

Table 1: Transmitter information parameters for the locations.

Key parameters		Cell identity	
		FC0125	FC3112
Technical parameters	Transmission power (dBm)	10	10
	Frequency (GHz)	3.5	3.5
	Transmitter height (m)	35	30
	Coverage area (km)	0.71	0.69
	Cell radius (mm)	1.2	2.4
	Modulation types	16-QAM	64-QAM
	Channel coding techniques	LDPC	LDPC
	Duplexing method	FDD	FDD
	Multiple access method	CDMA	CDMA
	Handover method	Soft handover	Soft handover
	Antenna type	Sectorial	Sectorial
	Antenna gain (dBi)	20	20
	Number of antenna ports	16 ports	16 ports
Geographic parameters	Environment	Regular urban	Irregular urban
	Latitude (°)	9.0853	9.0736
	Longitude (°)	7.4724	7.4950

Figure 4: Measurement setup.

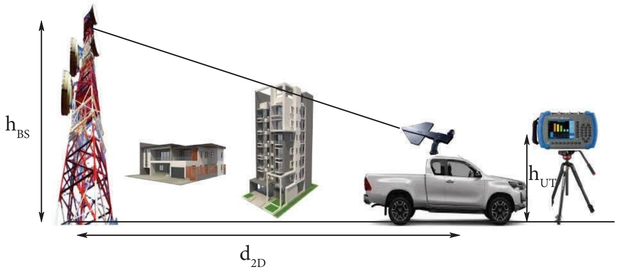


Figure 5: Illustration of the measurement campaign.

area under consideration. The environment was then set up by specifying the terrain model and identifying structures such as buildings, trees, and other objects that could obstruct radio wave propagation. Following this, transmission parameters were established, including the frequency, power, and antenna characteristics for both the transmitting and receiving devices. Appropriate propagation models were then selected to suit the scenarios under consideration. Subsequently, a ray tracing simulation was performed for the defined study area using the specified transmission parameters. Upon completion of the simulation, path loss data were collected. This dataset, generated by the RF planning tool software, included detailed information about the path loss at each specific location, taking into account the effects of the transmitted signal and the surrounding environment.

4.2.3. Geographical Data. To obtain geographical data for path loss analysis, the PL5 software tool was used to extract crucial geographical information about the study area, including path profile, terrain elevation, path length, and path inclination. The process began by launching the PL5 software and importing the geographical map data of the study area from OpenStreetMap. The study area was then defined within these map data, selecting the specific region for path loss analysis. Transmitter locations were identified on the map within the selected study area, and relevant transmitter information, such as frequency, power, and antenna height, was entered. Receiver locations were specified next, either as predetermined points or using a grid-based approach to cover the entire study area. The path loss calculation process was then initiated. Upon completion, the relevant geographical data were exported into a suitable format for integration with other tools.

4.2.4. Weather Station. Generating a path loss dataset from a weather station involves considering weather-related parameters that can influence signal propagation. In this case, relevant weather parameters that are known to impact signal propagation, such as temperature and humidity, were collected from the Nigerian Meteorological Agency (NiMet) at regular intervals for a period of two years, while considering factors such as environmental changes and local variations.

4.2.5. Dataset Description. The complete dataset used for this study comprised a total of 280 samples collected from both measurement campaigns and complementary sources. From the field measurements, 71 samples were recorded along LoS paths and 71 along NLoS paths in the regular urban environment (FC0125), whereas 69 LoS samples and 69 NLoS samples were collected in the irregular urban environment (FC3112). This resulted in 280 empirical path loss measurements across distances ranging from 10 to 710 m in FC0125 and 10–690 m in FC3112, with 10-m intervals. Measurements were repeated three times at each point, and averages were used to improve accuracy.

The equipment setup included a stationary 5G base station (Tx) at 3.5 GHz and a handheld spectrum analyzer (Agilent N9344C) connected to an HE200 directional antenna mounted at 1.5 m on a moving vehicle, serving as the receiver (Rx). Weather-related features (temperature and humidity) were collected from the NiMet over a 2-year period (2021–2022) at regular intervals. In addition, 3D ray tracing simulations using Altair WinProp and geographical features extracted via PL5 software and OpenStreetMap were incorporated, generating a total of 18 candidate features for model training.

4.3. Model Development and Evaluation

4.3.1. Optimization of the COST-231 Hata Model. To optimize the COST-231 Hata model for 3.5-GHz operations, the parameters of the conventional COST-231 Hata model were fine-tuned. This involved an initial adjustment of the model's parameters to better fit the specific dataset. This preliminary step aimed to reduce discrepancies between the model's predictions and the actual measurements. Following this, the parameters were further optimized using the "*scipy.optimize.minimize*" module in Python, which minimizes a scalar function of one or more variables using a solver. In this optimization, the Constrained Optimization by Linear Approximations (COBYLA) solver was employed. This solver was selected due to its superior performance based on both R^2 score and MSE compared to other solvers that were tested, which include Nelder-Mead, Powell, L-BFGS, TNC, SLSQP, and trust-constr. The algorithm that was used to realize the optimized COST-231 Hata model is presented in Algorithm 1.

4.3.2. Training of the RF Model. To train the RF model based on the RFE technique, the generated dataset from the 3D ray tracing, weather station, and geographic data was combined with that of the measurement campaign to construct a dataset of eighteen for the training and testing of the RF model.

The input features include the path length between the Tx and Rx, receiver (Rx) coordinates, elevation, relative humidity, temperature, path inclination, true azimuth, received signal strength of all the locations, and the features considered in the conventional COST-231 Hata model. The flowchart for training and evaluating the performance of the RF model is presented in Figure 6.

After applying the RFE technique, eleven out of the eighteen significant features were identified as the most crucial contributors to path loss and retained, including environmental and geographical attributes that enhanced dataset diversity. This combination of empirical measurements, ray tracing simulations, geographical information, and long-term weather data ensured that the dataset captures variations across urban morphology, LoS/NLoS propagation, environmental conditions, and measurement scenarios, thereby improving the robustness and generalizability of the developed hybrid model. Figure 7 illustrates the importance of these features after the implementation of the RFE technique.

```
# Given equation
def path_loss_equation (constants, X):
  f = X ['Frequency (MHz)']
  ht_x = X ['Tx_height']
  hr_x = X ['Rx_height']
  d = X ['Distance (km)']
  C_m = 3 # dB for urban environments
  a_hr_x = (1.1 * np.log10 (f) − 0.7) * hr_x − (1.56 * np.log10 (f) − 0.8)
  PL = (
    constants [0] +
    constants [1] * np.log10 (f) −
    constants [2] * np.log10 (ht_x) +
    C_m −
    a_hr_x +
     (constants [3] − constants [4] * np.log10 (hr_x)) * np.log10 (d)
  )
  return PL
# Mean squared error between calculated PL and actual PL from data
def loss_function (constants, X, y):
  calculated_PL = path_loss_equation(constants, X)
  actual_PL = y # Actual path loss from the dataset
  mse = np.mean((calculated_PL − actual_PL)**2)
  return mse
# Initial guess for constants
initial_constants = [46.33, 33.9, 13.82, 44.99, 6.55]
# Optimize the constants
# bounds = [(−345, −280), (−345, −280), (−345, −280), (−345, −280), (−345, −280)]
optimized_constants = minimize (loss_function, initial_constants,
          args = (X_train, y_train),
          method = 'COBYLA')#, bounds = bounds)
print ("Optimized Constants:", optimized_constants.x)
```

Algorithm 1: Optimized COST-231 Hata model.

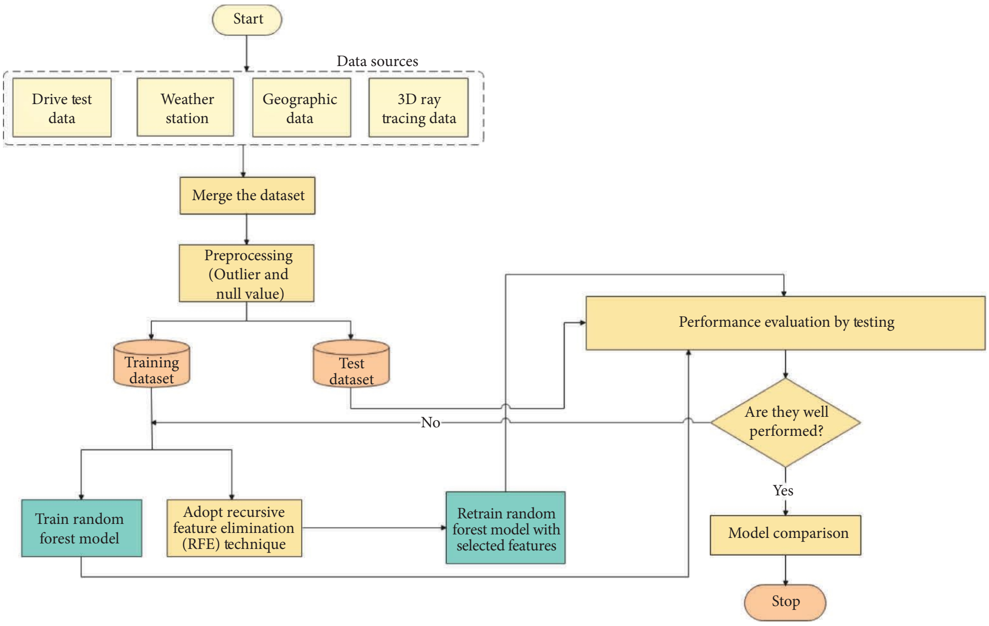


Figure 6: Adopted flowchart for training and evaluating the performance of the RF model.

To evaluate the effectiveness of the feature selection process, a comparison was made between the RF model trained with all 18 features and the RF model after applying RFE, which retained 11 features, as presented in Table 2. The results indicate that RFE led to a consistent improvement in predictive accuracy, reducing RMSE from 7.21 to 6.42 dB

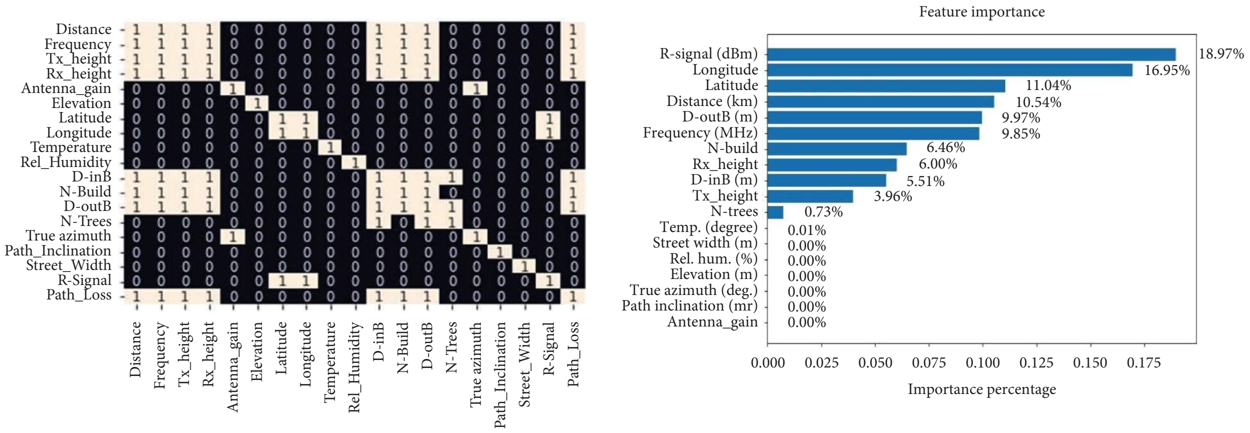


Figure 7: Feature importance heat map.

Table 2: Performance comparison of the random forest model before and after RFE.

Model variant	RMSE (dB)	MAE (dB)	MAPE (%)	R^2
RF (all 18 features, no RFE)	7.21	6.74	9.12	0.69
RF (after RFE, 11 features)	6.42	5.88	8.15	0.74

and mean absolute error (MAE) from 6.74 to 5.88 dB, while also improving R^2 from 0.69 to 0.74. This demonstrates that RFE successfully eliminated redundant or less informative predictors, thereby enhancing model generalization and reducing overfitting. Importantly, the performance gain achieved with fewer features also highlights the efficiency of the hybrid model framework, as the optimized RF component feeding into the empirical model is both more accurate and computationally efficient.

4.3.3. Development of the Hybrid Model. To develop the hybrid path loss prediction model, the RFE-based improved RF model was combined with the optimized COST-231 Hata model using their weighted average, as presented in Algorithm 2. To determine the optimal weighting for combining the RF model and the optimized COST-231 Hata model, several weighting ratios were empirically tested, including 50:50, 60:40, 70:30, and 80:20. The evaluation was based on standard performance metrics such as RMSE, MAE, and coefficient of determination (R^2). Among these, the 70:30 weighting (favoring the RF predictions) consistently yielded the lowest error values and the highest R^2, demonstrating a superior balance between capturing nonlinear feature interactions through RF and maintaining propagation trend fidelity from the COST-231 Hata model. Consequently, the 70:30 weighting was adopted as the most effective combination strategy for the proposed hybrid model. The flowchart for the proposed model is shown in Figure 8.

4.3.4. Comparative Analysis, Validation, and the Generalization of the Hybrid Model. To conduct a comparative analysis of the hybrid model, the prediction results of the proposed hybrid path loss model were compared with those of the optimized COST-231 Hata model and the RFE-based RF model, using performance metrics such as MAE, RMSE, and MAPE.

To validate the new model with independent path loss data, its path loss prediction capability was tested across diverse scenarios (irregular urban environments), as each environment presents unique challenges and characteristics that can significantly impact signal propagation. Testing the model in such an environment can assess its generalizability and robustness, confirming its ability to predict path loss under varying conditions accurately. This validation process helps mitigate the risk of relying on a model that may perform well in one environment but falter in others, thereby enhancing its credibility. Additionally, comparing predictions with actual measurements in different environments allows for an accurate assessment of the model's accuracy and identification of any limitations or discrepancies, ultimately contributing to more informed decision-making in system design and deployment efforts.

To generalize the applicability and effectiveness of the proposed hybrid path loss model, a validation was further conducted in an environment considered in [35]. The validation process involved applying the hybrid model to this new environment and comparing its predicted path loss values against the measured path loss obtained. The focus was on the LoS scenario to ensure consistency with the reference study.

5. Results and Discussion

5.1. Received Signal Strength. The results of the path loss measurements carried out in the regular urban environment are shown in Figure 9. The fluctuations experienced in the received signal strength between 500 and 580 m are primarily attributed to the mixture of LoS and NLoS conditions, characteristic of a Rician environment. A Rician environment occurs when the received signal consists of a dominant direct path combined with multipaths, leading to variations in signal strength.

These reflections cause multiple copies of the signal to arrive at the receiver with varying delays and phases, leading to constructive and destructive interference. Additionally,

```
BEGIN FUNCTION combine_predictions (COST-231_predictions, rf_predictions, rf_weight = 0.7):
  Combined_predictions = (rf_weight ∗ rf_predictions) + ((1−rf_weight) ∗ COST-231_predictions)
  | RETURN combined_predictions
END FUNCTION
# Get importance weight for RF model (sum of importance as a percentage)
Rf_weight = np.sum (feature_importances) ∗ 100
```

Algorithm 2: Hybrid model.

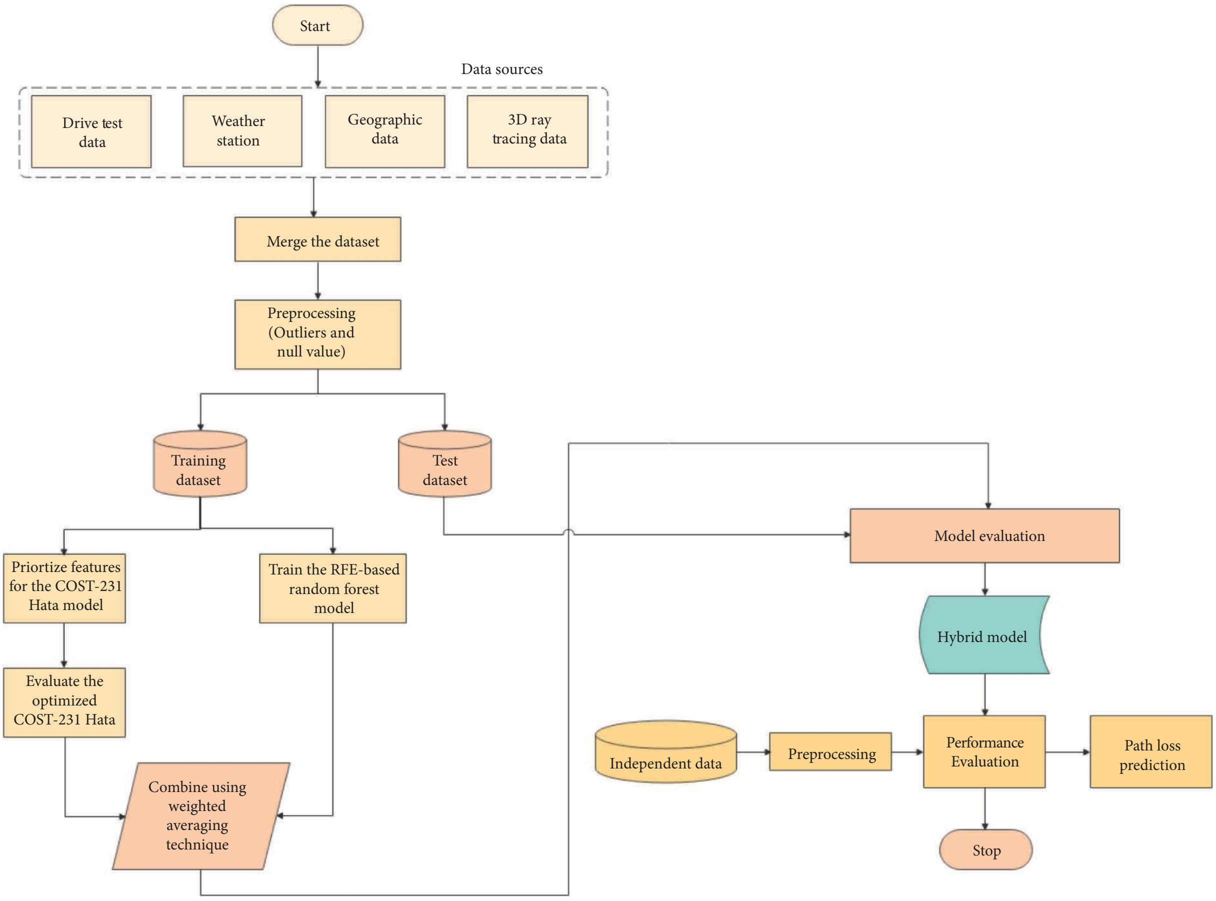


Figure 8: Flowchart of the proposed hybrid model.

the Fresnel zone effect is likely contributing to these fluctuations. The Fresnel zone is the area around the direct path between the transmitter and the receiver that must remain relatively clear of obstructions to minimize diffraction and interference. Given the transmitting antenna height of 30 m and the complex urban terrain, objects intruding into the Fresnel zone can cause further signal variability.

Similar fluctuations in RSS were observed in the NLoS scenario between 300 and 370 m. In NLoS conditions, the signal path is obstructed, forcing the signal to reflect and diffract around obstacles to reach the receiver. This results in more pronounced multipath effects compared to the LoS scenario, leading to the observed fluctuations.

An unexpected improvement in RSS was observed at 680 m, close to the cell edge at 710 m. There is a noticeable increase instead of a further reduction in signal strength. This phenomenon is likely due to signals the mobile receives from neighboring BSs, including the serving BS. Soft handovers are a common feature in code division multiple access (CDMA) systems, which the station under consideration uses. During a soft handover, a mobile device simultaneously communicates with multiple base stations to ensure a seamless transition from one cell to another. The improvement in signal strength at 680 m indicates that the mobile device began receiving signals from an adjacent base station, leading to a stronger combined signal as it transitioned between cells. This handover mechanism ensures continuous connectivity and enhanced signal quality, especially near the cell edge.

The results of these measurements in the irregular urban environment are presented in Figure 10. From 70 m up to 420 m away from the transmitter, the RSS along the LoS

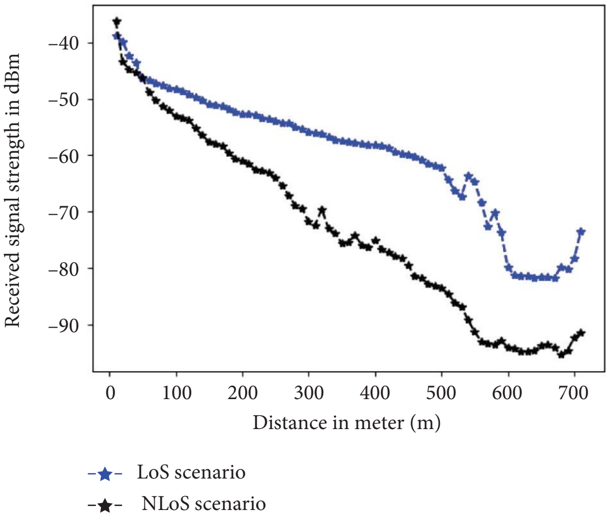


Figure 9: Average measured signal strength for the regular urban environment.

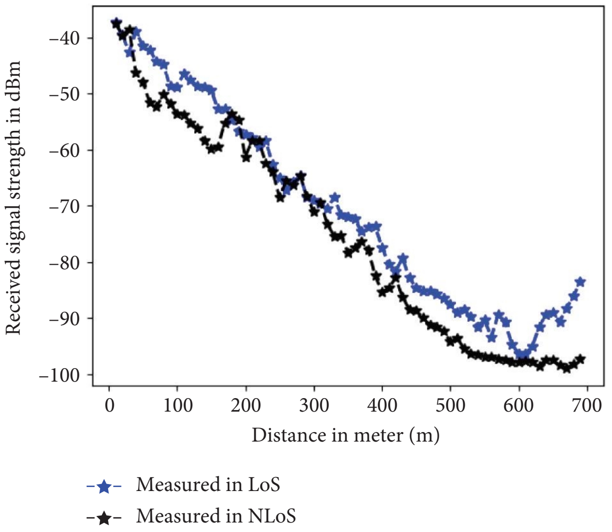


Figure 10: Average measured signal strength for the irregular urban environment.

scenario underperformed relative to expectations. This underperformance can be attributed to the presence of obstructions and reflective surfaces within the urban environment. Although the scenario is designated as LoS, the urban setting likely introduced elements of NLoS conditions. Structures and irregular street patterns may have caused signal reflections, diffraction, and scattering, creating a hybrid propagation environment. This mixed scenario aligns with the characteristics of Rician fading, where there is a predominant LoS path combined with multiple scattered paths. The variations and reduced signal strength suggest that the direct path was periodically obstructed and that the signal was influenced by multipath propagation.

Fluctuations in RSS were observed in both LoS and NLoS scenarios within the 70–420 m range. In the LoS scenario, these fluctuations point toward the presence of Rician fading, where the signal experiences interference from both the direct path and reflected paths. In areas where the direct path is intermittently blocked or heavily affected by reflections, Rayleigh fading might also be occurring, characterized by the absence of a dominant direct path and the presence of numerous scattered paths. The combination of these fading effects can cause significant variability in signal strength, as seen in the measurement results.

5.2. Optimized COST-231 Hata Model.

After tuning the parameters of the COST-231 Hata model to suit the dataset, using Algorithm 1 in Section 4.3.1, the optimized constants became 15.06035358, −40.27321111, −8.95983077, 84.0493291, and −39.38700645.

Therefore, the optimized COST-231 Hata model is presented as follows:

$$\begin{aligned}\mathrm{PL}_{\mathrm{opt}} &= 15.06 - 40.27\log_{10} f - 8.96\log_{10} ht_x + C_m \\ &\quad - a\left(hr_x\right) + \left(84.05 - 339.39\log_{10} hr_x\right)\log_{10} d,\end{aligned} \tag{7}$$

$$a\left(hr_x\right) = \left(1.1\log_{10} f - 0.7\right) \times hr_x - \left(1.56\log_{10} f - 0.8\right), \tag{8}$$

where $\mathrm{PL}_{\mathrm{opt}}$ is the path loss of the optimized COST-231 Hata model in dB, f is the frequency in MHz, d is the distance from the transmitting station to the receiving antenna in km, ht_x is the transmitter antenna height in m, hr_x is the receiver antenna height in m, and C_m is the correction factor, which maintains its value as 3 dB for urban environments.

5.3. Hybrid Path Loss Model.

As discussed in Section 4.3.3, the models were combined by calculating the weighted average of the two input predictions. The equations for the hybrid model are presented as follows:

$$\mathrm{RF}_{sf}^{w} = \sum_{i=0}^{n}\left(X_i^n\right), \tag{9}$$

$$\mathrm{PL}_H = \left(\mathrm{RF}_{sf}^{w} \times \mathrm{RF}_{sf}\right) + \left(\left(1 - \mathrm{RF}_{sf}^{w}\right) \times \mathrm{PL}_{\mathrm{opt}}\right), \tag{10}$$

where RF_{sf}^{w} is the weighted importance of the RF selected features X_i^n, RF_{sf} is the RF retrained model using the selected features, $\mathrm{PL}_{\mathrm{opt}}$ is the path loss of the optimized COST-231 Hata model, and PL_H is the path loss of the hybrid model.

5.4. Path Loss Prediction.

The graphical results in Figure 11, depicting the prediction of path loss in an urban environment at 3.5 GHz, offer compelling insights into the efficacy of different models. In the LoS scenario, the graphical representation showcases the superior performance of the proposed hybrid model compared to both the RF model and the optimized COST-231 Hata model in predicting path loss. Notably, although the RF model demonstrates

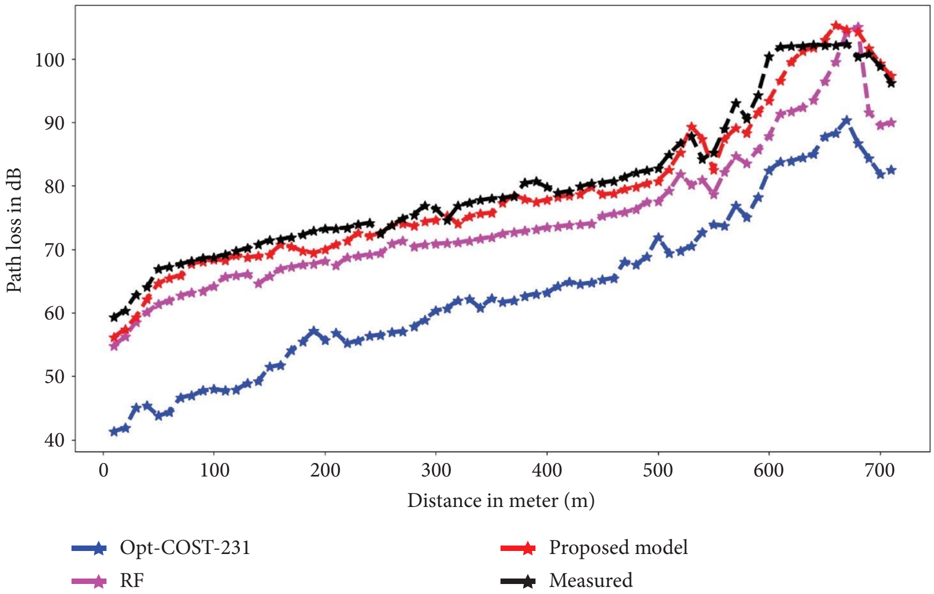


Figure 11: Path loss prediction in the regular urban environment in the LoS scenario.

commendable performance compared to the COST-231 Hata model, the proposed hybrid model surpasses both when compared to the measured path loss.

This outcome highlights the drawbacks of the optimized COST-231 Hata model, which notably underperformed in the context of 3.5 GHz, even though it was fine-tuned and optimized. The graphical representation serves as compelling evidence, illustrating the inadequacy of the optimized COST-231 Hata model for predicting path loss at the higher frequency of 3.5 GHz, thereby emphasizing the necessity and effectiveness of the proposed hybrid model in accurately modeling path loss within the urban environment at this frequency range.

Conversely, the results, shown in Figure 12, depicting path loss prediction in the NLoS scenario within an urban environment at 3.5 GHz, provide valuable insights into model performance. In the range of 10 m–710 m, the proposed hybrid model demonstrates superior predictive capabilities compared to the RF model and the optimized COST-231 Hata model.

However, within the first 350 m, an interesting trend emerges: The proposed hybrid model exhibits underperformance relative to the later stages of the measurement range. This underperformance, while notable, is considerably less pronounced when compared to the RF and optimized COST-231 Hata models within the same distance range. Beyond the initial 350 m, the hybrid model showcases a marked improvement, outperforming both the RF and optimized COST-231 Hata models, highlighting its ability to better predict path loss in the NLoS scenario at 3.5 GHz within an urban environment. This performance pattern underscores the hybrid model's advantage in accurately modeling path loss, especially when compared to the limitations observed in the other models in the NLoS scenario.

Based on the performance metrics, the optimized COST-231 Hata model exhibits RMSE values of 17.23 dB in the LoS scenario, whereas in the NLoS scenario, the RMSE value spikes to 26.19 dB, showcasing higher error rates and notably elevated MAE values in both scenarios, as shown in Figure 13. The RF model demonstrates improved performance with RMSE values of 5.69 dB (LoS) and 8.88 dB (NLoS).

However, the standout performer emerges as the proposed hybrid model, showcasing remarkable accuracy with RMSE values of 2.05 dB (LoS) and 3.31 dB (NLoS). These results substantiate the superior predictive prowess of the proposed model, significantly outperforming both the optimized COST-231 Hata and RF models in path loss prediction within this regular urban environment. The substantial reduction in RMSE values, coupled with lower MAE values across both scenarios, underscores the excellence and precision of the proposed model, affirming its efficacy and reliability in accurately modeling path loss in this specific urban environment.

5.5. Validation of the Hybrid Model.

5.5. Validation of the Hybrid Model. The results for the validation are hereby presented in Figure 14, illustrating path loss prediction at 3.5 GHz in the irregular urban environment for the purpose of testing the stability of the model and generalizing the applicability of the proposed hybrid model in urban environments.

Across the span of 10–490 m, the proposed model demonstrates a strong alignment with the measured path loss data, showcasing its accuracy and reliability in modeling signal propagation within this complex urban landscape. However, beyond 490 m, a noteworthy trend emerges: The proposed model starts to exhibit an overprediction of the path loss. This deviation, though limited to the latter part of the measurement range, contrasts with its prior accuracy. In contrast, while demonstrating similar trends, the RF and optimized COST-231 Hata models display less consistency and accuracy in predicting path loss within this heterogeneous urban environment. The proposed model's proficient

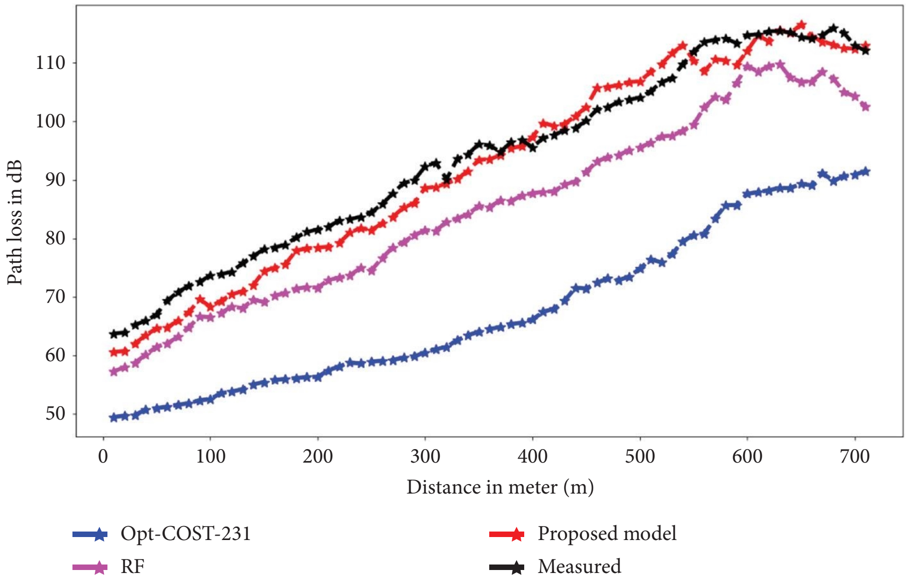


Figure 12: Path loss prediction in the regular urban environment in the NLoS scenario.

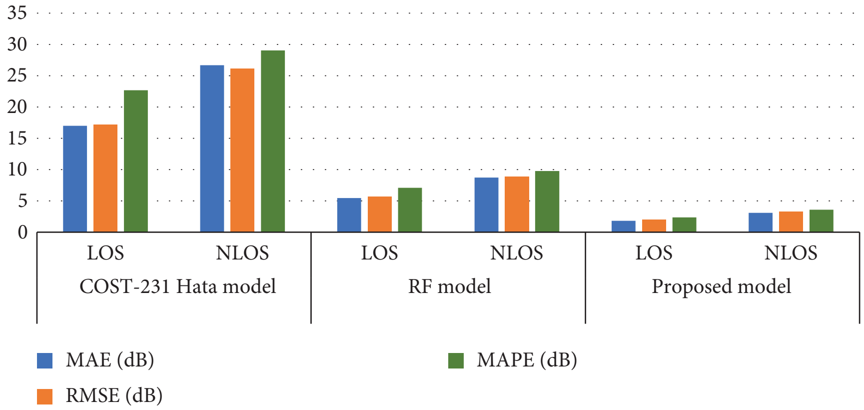


Figure 13: Models' prediction performances in the regular urban environment.

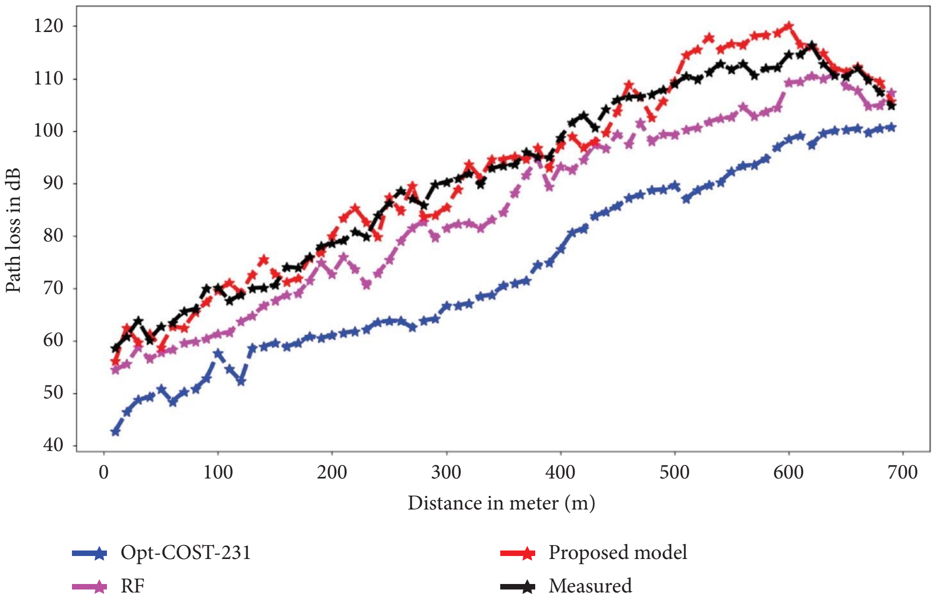


Figure 14: Path loss prediction in the irregular urban environment in the LoS scenario.

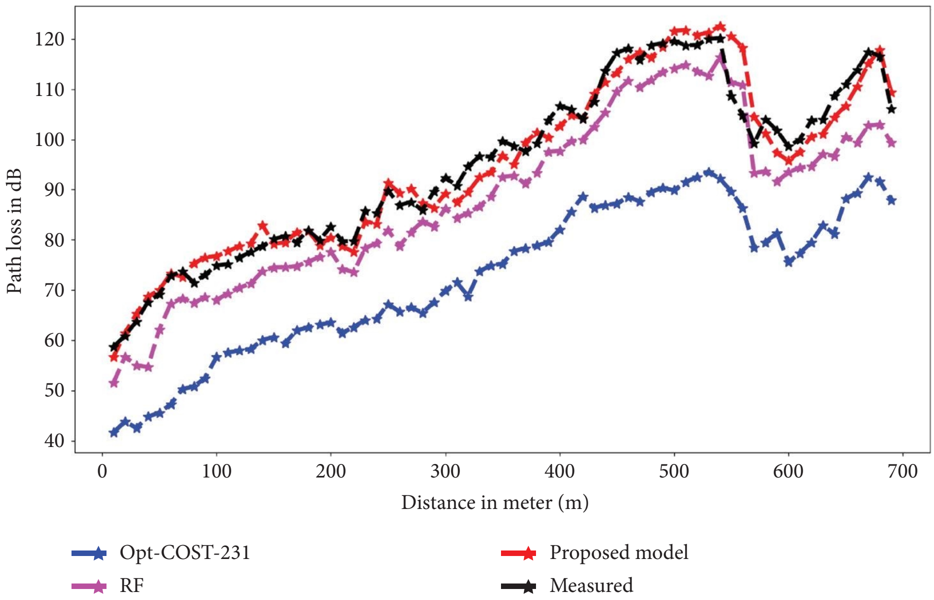


Figure 15: Path loss prediction in the irregular urban environment in the NLoS scenario.

alignment with measured path loss up to 490 m underscores its suitability for capturing the intricate dynamics of signal propagation in the LoS scenario. Nonetheless, the observed overprediction beyond 490 m suggests a need for further refinement or adjustment in modeling the path loss dynamics in this specific context at 3.5 GHz.

Similarly, in the NLoS scenario, as shown in Figure 15, for the entire range considered, up to 710 m, the proposed model showcases a strong alignment with the measured path loss data. This robust consistency underscores the model's accuracy and reliability in predicting path loss in this complex urban setting. In contrast, the RF and optimized COST-231 Hata models exhibit notable deviations and inaccuracies compared to the measured path loss.

The proposed model's ability to consistently fit the measured path loss across the range of up to 710 m in the NLoS scenario highlights its superior performance and suitability for accurately modeling path loss dynamics in this diverse and less structured urban environment at 3.5 GHz.

The validation results for path loss prediction models in an irregular urban environment, evaluated using MAE, MAPE, and RMSE, reveal the comparative performance of the models under scrutiny, as presented in Figure 16. The optimized COST-231 Hata model exhibits considerable discrepancies in accuracy within this irregular urban environment. Notably, high RMSE, MAE, and MAPE values for both LoS and NLoS scenarios indicate a lack of precision in capturing the complex signal propagation dynamics within this environment.

Conversely, the RF model showcases improved accuracy compared to COST-231 Hata, presenting lower RMSE, MAE, and MAPE values in both LoS and NLoS scenarios. Despite this improvement, the standout performer is undoubtedly the proposed hybrid model, which consistently outperforms both the optimized COST-231 Hata and RF models across all metrics in both scenarios. With substantially lower RMSE, MAE, and MAPE values, the proposed model demonstrates exceptional accuracy and reliability in predicting path loss within this irregular urban environment at 3.5 GHz.

The results of validating the proposed hybrid path loss model using data from [35] are presented in Figure 17 and Table 2. Figure 17 shows the path loss prediction in a comparative environment, assessing the performance of the proposed hybrid model, the optimized COST-231 Hata model, and the RF model. This validation serves to evaluate how closely each model aligns with the measured path loss values across various distances.

The results reveal that, over the entire range considered (up to 800 m), the proposed hybrid model consistently outperforms the other models, exhibiting a closer alignment with the measured path loss. This supports earlier findings, confirming that the hybrid model provides more accurate predictions in both regular and irregular urban environments compared to the other models. The RF model, while performing reasonably well, demonstrated some degree of underperformance, though it was not as pronounced as that of the optimized COST-231 Hata model. The optimized COST-231 Hata model significantly overpredicted path loss beyond 150 m, only beginning to align with the measured values around 430 m.

The validation conducted with the independent dataset from Ogun State at 1.6 GHz was not intended as a direct same-band evaluation but rather as a cross-band validation exercise to examine the generalization capability of the proposed hybrid model beyond its training frequency of 3.5 GHz. Although the model was optimized for midband operation, the 1.6-GHz results are valuable in showing that the hybrid model consistently outperforms the benchmark models even outside its training band. This suggests that the proposed approach is adaptable and robust across different frequency ranges. Nevertheless, it is acknowledged that the

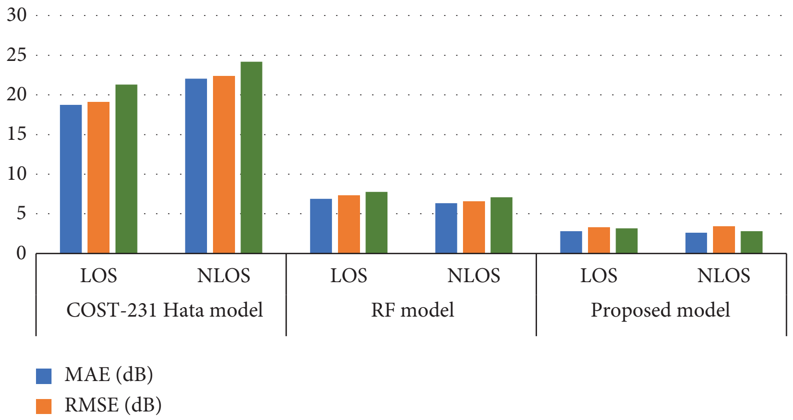


FIGURE 16: Models' prediction performances in the irregular urban environment.

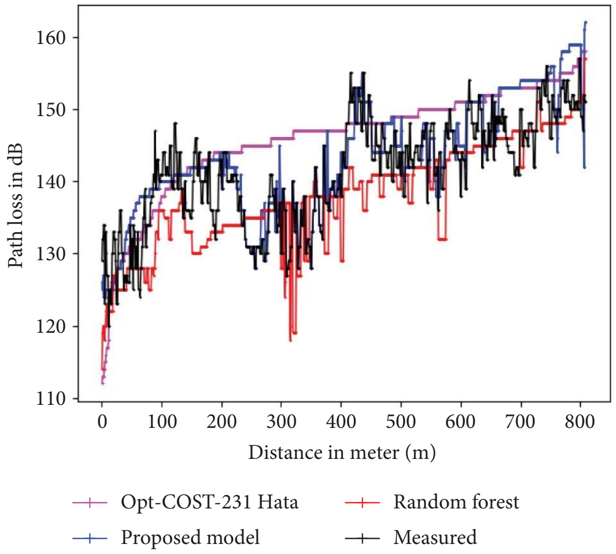


FIGURE 17: Path loss predictions in comparative environment.

TABLE 3: Comparison of the examined models.

Accuracy metrics	Examined models		
	Opt-COST-231 Hata	Random forest (RF)	Proposed hybrid model
MAE (dB)	5.3	3.1	2.8
RMSE (dB)	8.4	3.4	2.1
MAPE (%)	4.9	1.6	1.2

model could achieve even better predictive accuracy if retrained or fine-tuned specifically for 1.6 GHz, which is recommended as future work.

To reveal the individual performance of the examined path loss models, Table 3 presents a comparative analysis using three adopted performance metrics: MAE, RMSE, and MAPE. The results demonstrate that the proposed hybrid model outperforms both the RF and optimized COST-231 Hata models, achieving the lowest MAE of 2.8 dB, RMSE of 2.1 dB, and MAPE of 1.2%. This superior performance highlights the robustness of the hybrid model and underscores its potential for broader applicability in diverse urban environments.

To further assess the generalizability of the proposed hybrid model, a k-fold cross-validation procedure was applied to the training dataset in both LoS and NLoS scenarios across regular and irregular urban environments. The averaged error metrics (RMSE, MAE, and MAPE) obtained from the folds demonstrated close agreement with the initially reported results in Figures 13 and 16, thereby confirming the stability and robustness of the hybrid model. This consistency across multiple data partitions highlights

Table 4: Cross-validation results of the hybrid model in LoS and NLoS scenarios.

Scenarios	Environment	RMSE (dB)	MAE (dB)	MAPE (%)
LoS	Regular urban	2.05	1.82	2.40
LoS	Irregular urban	3.31	2.82	3.18
NLoS	Regular urban	3.34	3.12	3.62
NLoS	Irregular urban	3.43	2.63	2.81

the model's reliability and applicability for practical deployment in diverse urban settings. Table 4 presents the cross-validation results of the hybrid model in LoS and NLoS scenarios.

The cross-validation results in Table 4 confirm the stability and generalizability of the hybrid model across LoS and NLoS scenarios in both regular and irregular urban environments.

6. Conclusion

This paper focused on developing a hybrid path loss model for 5G communication networks in the midband frequencies, the strengths of machine learning, and empirical modes. In particular, the strengths of an optimized COST-231 Hata model and the RFE technique-based RF model were combined in a 70:30 ratio to form the hybrid. The results demonstrated a significant improvement in the RF model's performance when employing RFE, leading to a 30.52% reduction in MSE and an 8.08% decrease in MAE. The hybrid model exhibited superior predictive capabilities in both LoS and NLoS scenarios at 3.5 GHz for the two distinct urban environments. In the LoS scenario, the hybrid model surpassed both the RF and COST-231 Hata models, emphasizing its accuracy and reliability, particularly within the critical distance range of 5–600 m. In the NLoS scenario, the hybrid model consistently outperformed the RF and COST-231 Hata models across the entire range, demonstrating its suitability for accurately modeling path loss dynamics in diverse and less structured urban environments at 3.5 GHz.

Although challenges were observed, such as the overprediction trend in the LoS scenario of the irregular urban environment, the proposed hybrid model showcased remarkable consistency and accuracy, highlighting its potential for enhancing the precision of path loss predictions in 5G communication networks. In addition, a dataset from a completely different urban environment was used to validate the hybrid model, and it performed considerably well. Further refinements are suggested to address specific deviations, ensuring the model's robustness and applicability in various deployment scenarios. Overall, the study contributes valuable insights into advancing path loss modeling for the optimization of 5G communication networks.

Data Availability Statement

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

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

This study was financially supported by the Council for Scientific and Industrial Research, South Africa.

Acknowledgments

The authors thank LiviaSoft Technologies Ltd. for providing the RF software tools. They also thank the Nigerian Communications Commission (NCC) for supplying the hardware equipment used in the measurement campaigns.

References

[1] M. Singh and K. Weidner, "Types and Performance of High Performing Multi-Mode Polymer Waveguides for Optical," in *Optical Interconnects for Data Centers*, ed. T. Tekin, R. Pitwon, A. Hakansson, and N. Pleros (Woodhead Publishing, 2017), 157–170.

[2] F. E. Shaibu, E. N. Onwuka, N. Salawu, and S. S. Oyewobi, "Performance Analysis of Path Loss Models for Wireless Communications at 3.5 GHz and 23 GHz in a Regular Urban Environment," in *Presented at the 2nd International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS)* (Abuja, Nigeria, 2023).

[3] J. Feher, "Speed of Light," *ScienceDirect* (2023): https://www.sciencedirect.com/topics/engineering/speed-of-light.

[4] S. Salous, D. Gaillot, and J.-M. Molina, "IRACON Channel Measurements and Models," *Inclusive Radio Communications for 5G and Beyond* (New York: Elsevier Ltd, 2021), 49–105.

[5] K. Aziding, C. W. Byeon, M. S. Park, D. O. Cho, and Y. C. Lee, "Impact of Reflected Energy from Surrounding Environments on BEL over the Cluttered Path at 3, 6, and 10 GHz," in *Presented at the 2021 15th European Conference on Antennas and Propagation (EuCAP), Dusseldorf, Germany* (2021).

[6] K. Pokorny, *Top 10 Stories of 2022 Cover a Wide Range of Topics* (Oregon State University, 2023), https://today.oregonstate.edu/news/top-10-stories-2022-cover-wide-range-topics.

[7] B. Myagmardulam, N. Tadachika, K. Takahashi, et al., "Path Loss Prediction Model Development in a Mountainous Forest Environment," *IEEE Open Journal of the Communications Society* 2 (2021): 2494–2501, https://doi.org/10.1109/OJCOMS.2021.3122286.

[8] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion," *Applied Sciences* 9, no. 9 (2019): 1908, https://doi.org/10.3390/app9091908.

[9] M. A. Alim, M. M. Rahman, M. M. Hossain, and A. Al-Nahid, "Analysis of Large-Scale Propagation Models for Mobile Communications in Urban Area," *International Journal of Computer Science and Information Security* 7, no. 1 (2010): 135–139.

[10] A. Akinbolati and M. O. Ajewole, "Investigation of Path Loss and Modeling for Digital Terrestrial Television over Nigeria," *Heliyon* 6, no. 6 (2020): e04101, https://doi.org/10.1016/j.heliyon.2020.e04101.

[11] Z. Nadir, Z. Elfadhil, and F. Touati, "Path Loss Determination Using Okumura-Hata Model and Spline Interpolation for

Missing Data for Oman," in *Proceedings of the World Congress on Engineering*, 1 (London, UK, 2008), 1–6.

[12] N. S. Nkordeh, A. A. Atayero, and F. E. Idachaba, "LTE Network Planning Using the Hatta-Okumura and the COST-231 Hata Path Loss Models," in *World Congress on Engineering, WCE 2014*, 1 (London, UK, 2014), 1–5.

[13] T. J. Rouphael, "High-Level Requirements and Link Budget Analysis," in *RF and Digital Signal Processing for Software-Defined Radio*, ed. T. J. Rouphael (Elsevier Ltd, 2009), 87–122.

[14] H. Tataria, K. Haneda, A. F. Molisch, M. Shafi, and F. Tufvesson, "Standardization of Propagation Models for Terrestrial Cellular Systems: a Historical Perspective," *International Journal of Wireless Information Networks* 28, no. 1 (2021): 20–44, https://doi.org/10.1007/s10776-020-00500-9.

[15] M. Nadeem, M. Wasid, M. Nadeem, and S. Shahab, "Performance Comparison of Randomized and Non-Randomized Learning Algorithms Based Recommender," *International Journal of Next Generation Computing* 13, no. 3 (2022): 560–577, https://doi.org/10.47164/ijngc.v13i3.820.

[16] S. Maddali, "What are the Most Important Preprocessing Steps in Machine Learning and Data Science," *Data Science* (2023): https://towardsdatascience.com/what-are-the-most-important-preprocessing-steps-in-machine-learning-and-data-science-a7606d18f32a.

[17] P. Wang and H. Lee, "Meta-Learning Approaches for Indoor Path Loss Modeling of 5G Communications in Smart Factories," *ICT Express* 8, no. 2 (2022): 290–295, https://doi.org/10.1016/j.icte.2022.01.003.

[18] R. Meltzer, "What is Random Forest," *CAREERFOUNDRY* (2023): https://careerfoundry.com/en/blog/data-analytics/what-is-random-forest/.

[19] D. Nelson, "What is a KNN (K-Nearest Neighbors)," (2023), https://www.unite.ai/what-is-k-nearest-neighbors/.

[20] V. Chidambaram, *Algorithms Every Data Scientist Should Know* (STEM), https://www.codingal.com/coding-for-kids/blog/algorithms-every-data-scientist-should-know/.

[21] R. He, Y. Gong, W. Bai, and Y. Li, "Random Forest Based Path Loss Prediction in Mobile Communication Systems," in *Presented at the 2020 IEEE 6th International Conference on Computer and Communications (ICCC)* (Chengdu, China, 2020).

[22] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine-Learning Based Methods for Path Loss Prediction in Urban Environment for LTE Networks," in *Presented at the 16th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)* (Thessaloniki, Greece, 2020).

[23] G. Vergos, S. P. Sotiroudis, G. Athanasiadou, G. V. Tsoulos, and S. K. Goudos, "Comparing Machine Learning Methods for Air-to-Ground Path Loss Prediction," in *Presented at the 10th International Conference on Modern Circuits and Systems Technologies (MOCAST)* (Thessaloniki, Greece, 2021).

[24] K. Ahmad and S. Hussain, "Machine Learning Approaches for Radio Propagation Modeling in Urban Vehicular Channels," *IEEE Access* 10 (2022): 113690–113698, https://doi.org/10.1109/ACCESS.2022.3218622.

[25] S. Sukemi, A. F. Oklilas, M. W. Fadli, and B. Alfaresi, "Path Loss Prediction Accuracy Based on Random Forest Algorithm in Palembang City Area," *Jurnal Nasional Teknik Elektro* 20, no. 1 (2023): 1–7, https://doi.org/10.25077/jnte.v12n1.1051.2023.

[26] A. L. d. C. Barcellos, J. C. Duarte, and A. C. Mendes, "Radiofrequency Signal Levels Prediction Using Machine Learning Models," *IEEE Latin America Transactions* 21, no. 2 (2023): 351–357, https://doi.org/10.1109/TLA.2023.10015229.

[27] R.-T. Juang, J.-Q. Lin, and H.-P. Lin, "Machine Learning-Based Path Loss Modeling in Urban Propagation Environments," in *Presented at the 30th Wireless and Optical Communications Conference (WOCC)* (Taipei, Taiwan, October 2021).

[28] T. T. Nguyen, N. Yoza-Mitsuishi, and R. Caromi, "Deep Learning for Path Loss Prediction at 7 GHz in Urban Environment," *IEEE Access* 11 (2023): 33498–33508, https://doi.org/10.1109/ACCESS.2023.3264230.

[29] F. E. Shaibu, E. N. Onwuka, N. Salawu, S. S. Oyewobi, K. Djouani, and A. M. Abu-Mahfouz, "Performance of Path Loss Models Over Mid-Band and High-Band Channels for 5G Communication Networks: A Review," *Future Internet* 15, no. 11 (2023): 362–32, https://doi.org/10.3390/fi15110362.

[30] S. Aldossari and K.-C. Chen, "Predicting the Path Loss of Wireless Channel Models Using Machine Learning Techniques in mmWave Urban Communication," *Presented at the 2019 22nd International Symposium on Wireless Personal Multimedia Communications, Lisbon, Portugal* (2019).

[31] L. Juan-Llacer, "Path Loss Measurements and Modelling in a Citrus Plantation in the 1800 MHz, 3.5 GHz and 28 GHz in LoS," in *Presented at the 2022 16th European Conference on Antennas and Propagation (EuCAP)* (Madrid, Spain, 2022).

[32] E. M. D. Djomadji, K. I. Basile, F. S. Thierry, and T Emanuel, "COST 231-Hata Propagation Model Optimization in 1800 MHz Band Based on Magnetic Optimization Algorithm: Application to the City of Limbé," *Journal of Computer and Communications* 11, no. 02 (2023): 57–74, https://doi.org/10.4236/jcc.2023.112005.

[33] Z. Shakir, A. Al-Thaedan, R. Alsabah, M. Salah, A. Alsabbagh, and J. Zec, "Performance Analysis for a Suitable Propagation Model in Outdoor with 2.5 GHz Band," *Bulletin of Electrical Engineering and Informatics* 12, no. 3 (2023): 1478–1485, https://doi.org/10.11591/eei.v12i3.5006.

[34] F. E. Shaibu, E. N. Onwuka, N. Salawu, and S. S. Oyewobi, "Evaluating the Effectiveness of Machine Learning Models for Path Loss Prediction at 3.5 GHz With Focus on Feature Prioritization," *Nigerian Journal of Technology* 43, no. 4 (2025): 754–762, https://doi.org/10.4314/njt.v43i4.15.

[35] S. I. Popoola, A. A. Atayero, and O. A. Popoola, "Comparative Assessment of Data Obtained Using Empirical Models for Path Loss Predictions in a University Campus Environment," *Data in Brief* 18 (2018): 380–393, https://doi.org/10.1016/j.dib.2018.03.040.