Performance analysis of artificial neural networks for predicting propagation losses in suburban environments for 4G LTE and 5G networks

Bruno Jácome Cavalcanti[1]*, Gustavo Araújo Cavalcante[2], Laércio Martins de Mendonça[3]

[1] bruno.cavalcanti@ifpb.edu.br, [2] gustavo.cavalcante@ifpb.edu.br. Instituto Federal da Paraíba (IFPB), João Pessoa, Paraíba, Brasil
[3] laerciomartins.1753@gmail.com. Universidade Federal do Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte, Brasil
* Correponding author

Abstract
This study analyzes two distinct approaches for predicting path loss at frequencies of 800 MHz, 1800 MHz, and 2600 MHz in suburban areas. These frequencies are commonly utilized in broadcasting, 4G LTE, and 5G networks. Two models of Artificial Neural Networks (ANN) were implemented: an Error-Based Neural Network (EBNN), which incorporates error correction by combining empirical propagation models with an ANN, and a Terrain Parameters Based Neural Network (TBNN), which uses input parameters commonly applied in related studies, such as the distance from the transmitter to the receiver, receiver altitude, average terrain level, and azimuth angle between the transmitter and receiver. The performance of these models was evaluated using root mean square error (RMSE) and the Wilcoxon rank-sum test, comparing them with empirical propagation models such as SUI, ECC-33, Ericsson, and TR 25.942. The results were then compared with data obtained from a measurement campaign conducted along three routes in the city of Natal, Brazil. The findings from both simulations and actual measurements showed good metric alignment, particularly highlighting the performance of the error-based model. The primary contribution of this study is demonstrating that these techniques enable more accurate prediction of signal information, thereby reducing errors in the planning and implementation of wireless networks.
Keywords: 4G LTE; 5G networks; artificial neural networks; path loss.

Análise de desempenho de redes neurais artificiais para prever perdas de propagação em ambientes suburbanos para redes LTE 4G e 5G

Resumo
Este estudo analisa duas abordagens distintas para prever a perda de propagação em frequências de 800 MHz, 1800 MHz e 2600 MHz em áreas suburbanas. Essas frequências são comumente utilizadas em redes de transmissão de televisão, 4G LTE e 5G. Dois modelos de Redes Neurais Artificiais (RNA) foram implementados: uma Rede Neural Baseada em Erros (EBNN), que incorpora correção de erros combinando modelos de propagação empíricos com uma RNA, e uma Rede Neural Baseada em Parâmetros de Terreno (TBNN), que usa parâmetros de entrada comumente aplicados em estudos relacionados, como a distância do transmissor ao receptor, altitude do receptor, nível médio do terreno e ângulo de azimute entre o transmissor e o receptor. O desempenho desses modelos foi avaliado usando o erro quadrático médio (RMSE) e o teste de soma de postos de Wilcoxon, comparando-os com modelos de propagação empíricos como SUI, ECC-33, Ericsson e TR 25.942. Os resultados foram então comparados com dados obtidos de uma campanha de medição conduzida ao longo de três rotas na cidade de Natal, Brasil. Os achados de simulações e medições reais mostraram bom alinhamento métrico, destacando particularmente o desempenho do modelo baseado em erro. A principal contribuição deste estudo é demonstrar que essas técnicas permitem uma previsão mais precisa das informações do sinal, reduzindo assim os erros no planejamento e implementação de redes sem fio.
Palavras-chave: perda de propagação; redes 4G LTE; Redes 5G; redes neurais artificiais.

1 Introduction

Path loss prediction is a crucial factor in the planning and implementation of wireless networks. As a frequency signal propagates through various environments to reach its destination, its power diminishes gradually over distance. Accounting for path loss is therefore vital in the design of wireless link transmissions. Multiple mechanisms such as reflection, diffraction, absorption, scattering, and atmospheric conditions can lead to significant attenuation of radio signals (Saeed *et al.*, 2022).

To enhance the accuracy of communication systems planning, considerable efforts have been devoted to developing simulation methods for coverage prediction that closely estimate measured data. Various techniques have been employed to improve the efficiency of these simulation methods, minimizing errors and yielding more reliable outcomes. These methods aim to bridge the gap between theoretical models and real-world data by accurately estimating the performance of wireless networks. Advanced simulation techniques not only enable more precise coverage predictions but also help identify and reduce errors, resulting in more reliable and efficient network design and operation (Moraitis; Tsipi; Vouyioukas, 2020; Sung *et al.*, 2023).

By integrating various techniques – such as mathematical models, machine learning algorithms, and empirical data – researchers can refine simulations to better reflect real-world conditions. These advancements lead to more efficient networks, reduced costs associated with coverage misestimation, and ultimately, improved quality of service for end users. As communication requirements and technologies continue to evolve, it becomes imperative to develop robust modeling methodologies that adapt to new frequencies, ensuring accuracy and efficiency in network design and deployment (Sun *et al.*, 2016).

Artificial Neural Networks (ANNs) have seen substantial development in recent years, finding applications across a wide range of fields, including image and video compression (Ma *et al.*, 2019), big data analytics (Hernández *et al.*, 2020), autonomous vehicles (Ji *et al.*, 2018), meteorological forecasting (Rajendra *et al.*, 2019), and data assimilation (Cintra; Velho, 2012), among others. ANNs function by executing nonlinear mappings from a set of input values to a set of output values, which are processed through multiple layers of neurons. These input values are combined with respective synaptic weights within each layer to produce an output that is consistent with the given inputs (Ashrafijoo *et al.*, 2022).

Predicting path loss between two points can be framed as a problem involving the transformation of an input vector, which includes information such as the locations of the transmitter and receiver, frequency, and topographical and morphological details (Popescu; Nafornita; Constantinou, 2005). Given that ANNs are effective in solving nonlinear function approximation problems, they are well-suited for addressing the path loss prediction challenge.

Several related studies illustrate the potential of ANNs in this domain. For instance, Benmus, Abboud, and Shatter (2015) trained a neural network using measurement data from Tripoli, Libya, to predict signal strength across various frequencies. The authors compared their results with the Hata model, reporting Root Mean Squared Error (RMSE) values ranging from 4.32 dB to 6.73 dB in urban areas, 3.06 dB to 5.92 dB in suburban areas, and 4.07 dB to 4.88 dB in rural areas.

Eichie *et al.* (2017) conducted a comparative analysis of basic and ANN-based models for path loss prediction. They used propagation parameters such as the distance between transmitter and receiver, transmitting power, and terrain elevation as inputs to the neural network, with data collected along selected rural and suburban routes in Nigeria.

Jo *et al.* (2020) developed a path loss prediction approach based on machine learning techniques, including principal component analysis, artificial neural networks, and Gaussian processes. Their study collected data in a suburban environment in South Korea at various frequencies, achieving RMSE values of 7.87 dB, 8.96 dB, and 8.23 dB for different frequency bands using artificial neural networks.

In research by Popoola *et al.* (2021), propagation path loss in the Very High Frequency (VHF) band was characterized using various ANN learning algorithms and activation functions based on measurement data collected at 203.25 MHz in the urban environment of Ilorin, Nigeria. Among the ANN models tested the one using the hyperbolic tangent activation function (HTAF), the Levenberg-Marquardt training algorithm, and 80 neurons in the hidden layer yielded the best results, with an RMSE of 5.1 dB.

Popescu *et al.* (2002) presented results from applying a General Regression Neural Network (GRNN) to model path loss in urban and suburban areas. Various neural network models were tested for both environments, differing only in input parameters. The main inputs included distances between the transmitter and receiver, street width, building separation, and building height. Training data were collected in the city of Kavala and on the island of Santorini, Greece.

The GRNN-based model was compared with the Walfisch-Bertoni (WB) model and a modified version of the COST231-Walfisch-Ikegami (CWI) model. The neural network model demonstrated significant prediction improvement due to its generalization capabilities. RMSE values ranged from 5.35 dB to 8.66 dB in urban scenarios and from 3.68 dB to 5.23 dB in suburban scenarios. The study also introduced a hybrid error correction model based on combining the deterministic COST-Walfisch-Ikegami (CWI) model and a neural network, which was later expanded upon (Popescu; Nafornita; Constantinou, 2005).

The GRNN-based model, along with a Multilayer Perceptron Neural Network (MLP-NN) and a Radial Basis Function Network (RBFN), was developed using two network types: a simple neural network model with five inputs (distances between transmitter and receiver, street width, building height, building separation, and street orientation) and a hybrid neural network model based on error correction using the COST-Walfisch-Ikegami model.

The CWI model is considered a physical/statistical (or semi-empirical) model requiring terrain profile information, such as distances between transmitter and receiver, building heights, and spacing between buildings.

Ultimately, there was no significant difference between predictions made by the simple and hybrid versions. For urban environments, the simple RBF and MLP models achieved RMSE values of 5.35 dB and 6.55 dB, respectively, while the hybrid RBF and MLP models yielded RMSEs of 5.30 dB and 6.07 dB. In suburban areas, the hybrid RBF and MLP models achieved RMSEs of 3.71 dB and 3.77 dB, while the simple RBF and MLP models had RMSE values of 3.68 dB and 3.74 dB, respectively. This study was limited to covering only the frequency of 1870 MHz.

The work by Sanches and Cavalcante (2015) simplified this approach by developing a hybrid error-based model using empirical propagation models. This model only requires basic elements, such as the assigned frequency and distances between transmitter and receiver, to feed the network.

This study focused on a frequency of 900 MHz and employed the Okumura-Hata, Walfisch-Bertoni, Ibrahim-Parsons, Ericson, COST231-Hata, and Walfisch-Ikegami models, achieving RMSE values between 3.03 dB and 3.91 dB. The results, in terms of mean absolute error, standard deviation, and RMSE, confirmed the technique's good performance, yielding results closer to measurements.

This study proposes the implementation and application of two types of ANNs with similar architecture but different input types. In the first case, the network is inspired by the error-based hybrid propagation model developed by Popescu, Nafornita, and Constantinou (2005), with empirical models applied by Sanches and Cavalcante (2015). The designed neural network uses similar inputs: distances between transmitter and receiver and the error between measured data and predicted values by empirical models. This approach is referred to as the Error-Based Neural Network (EBNN).

The second approach is based on features of the terrain and the setup of measurements. The terrain/propagation characteristics used as inputs include the distance from the transmitter to the receiver, receiver altitude, average terrain level, and azimuth between transmitter and receiver. This approach is referred to as the Terrain Parameters Based Neural Network (TBNN).

In this study, the prediction models employed include the Ericsson 9999, Stanford University Interim (SUI), Electronic Communication Committee (ECC-33), and Technical Report (TR) 25.942 models, with their performance compared to models based on neural networks. These propagation models were selected due to their widespread use in telecommunications for predicting signal loss in various environments (3GPP, 2024a; Kumari *et al.*, 2011; Mathew; George; Pereira, 2017).

The Ericsson 9999 model is effective for both urban and suburban areas and is recognized for its accuracy, although it requires numerous input parameters, which may complicate its implementation. The SUI model, suitable for suburban environments and high frequencies, is relatively simple but less adaptable to specific conditions without additional adjustments.

The ECC-33 model is based on empirical measurements and is best suited for urban environments, offering high accuracy in densely populated areas. It can also be applied to suburban

areas, though it does not apply to rural settings. Finally, the TR 36.942 (2024c) model, part of the 3rd Generation Partnership Project (3GPP, 2024d) specifications for LTE and LTE-A networks, plans coverage and capacity across various environments, from rural to urban areas.

The experiment was conducted in suburban areas, focusing on the 800 MHz, 1800 MHz, and 2600 MHz frequencies. The SUI and ECC-33 models were applied to the 1800 MHz and 2600 MHz bands, while the Ericsson and TR 25.942 models covered all three frequency bands (GPP, 2024a). Table 1 summarizes the different frequency bands used for 4G LTE services, 5G, and broadcasting. Each frequency band is associated with multiple service categories, and the propagation models used in the study were applied to these frequency bands (3GPP, 2024b; 3GPP, 2024d).

Table 1 – Frequency bands covered in the study

Frequency band	5G	4G LTE	Broadcasting
800 MHZ	Bands: n20 (832 – 862 MHz/791 – 821 MHz) and n82 (832 – 862 MHz)	Bands: 20 (832 – 862 MHz/791–821 MHz), 28 (758 – 823 MHz) and 44 (703 – 803 MHz)	Present in digital dividend UHF band in Europe (UHF channels 61–69)
1800 MHz	Band n3 (1805 – 1880 MHz)	Band: 3 (1805 – 1880MHz/1710 - 1785 MHz) and 9 (1805– 1880MHz/1710- 1785 MHz);	–
2600 MHz	Bands n7 (2500 – 2570 MHz/2620 – 2690 MHz) and n38 (2570 – 2620 MHz)	Bands: 7 (2620 – 2690 MHz), 38 (2570 – 2620 MHz) and 69 (2570 – 2620 MHz)	–

Source: 3GPP (2024b, 2024d)

The methodology developed in this work is based on comparing these techniques to determine which provides simulation results that most closely align with the data obtained from measurements. A measurement campaign was conducted in three distinct routes within the district of Lagoa Nova, in Natal, Brazil.

The combination of neural networks with empirical models is a relatively uncommon approach in the existing literature, as empirical propagation models are more frequently utilized in comparative analyses against implemented neural networks.

MATLAB software version R2018a was employed to implement the computational methods. Data on distances, average terrain level, and altitude were obtained using SIGAnatel (Geographical Information System), a tool provided by Brazil's National Telecommunications Agency (ANATEL) for TV and FM broadcast channel feasibility simulations (Silva; Passos, 2007).

To evaluate the performance of each technique, two metrics were applied: the root mean square error (RMSE), which estimates the error difference in dB between the datasets, and the Wilcoxon rank-sum test, which provides an assessment of the similarity between the distribution of datasets.

The remainder of this paper is structured as follows: Section 2 details the measurement campaign; Section 3 discusses the propagation and ANN models; a comparative analysis between simulations and experimental data is presented in Section 4; and finally, Section 5 concludes the study and provides recommendations for future research.

2 Measurement campaign

Field measurements provide real empirical data on path loss and signal propagation, which are essential for validating and refining prediction models. Analyses are limited to theoretical models or simulations without measurement data, which may not accurately reflect actual environmental conditions. Measurement data also ensure the model can generalize well to new situations and varying conditions.

The measurement campaign was conducted in the Lagoa Nova district of Natal, Brazil. Measurements were taken along routes between the UFRN campus and nearby streets. During all campaigns, the weather was predominantly sunny, with clear skies and few clouds. The site featured regular vegetation density and medium-sized buildings, characterizing the environment as suburban.

A Continuous Wave (CW) signal was transmitted using a Rhode & Schwarz broadband amplifier (model R&SBBA150, 9 KHz – 6 GHz) and an Anritsu radio transmitter (model MG3700A, 50Hz – 6 GHz) with a power output of 15 watts. Two pairs of Pasternack directive antennas were employed for both transmission and reception (Pasternack, 2024): a panel antenna (2.5 GHz - 2.7 GHz) with a nominal gain of 14 dBi for the 2600 MHz band, and a dual-panel antenna (806-960 MHz and 1710-2500 MHz) with a nominal gain of 7 dBi for the 1800 MHz and 800 MHz bands.

For signal measurement, an Anritsu spectrum analyzer, model MS2721B, was utilized. This analyzer also featured an integrated Global Positioning System (GPS) device, which provided the precise location of the measured points.

The Rhode & Schwarz R&SBBA150 broadband amplifier covers a wide frequency range, from 9 KHz to 6 GHz, encompassing the entire frequency range of the transmitted signal. This broad coverage ensures that the amplifier can amplify signals across all relevant frequency bands without requiring changes in equipment. Combined with the transmitter's output power (15 watts) and the characteristics of the antennas used, the amplifier's power ensures that the transmitted signal is detectable during measurements.

The Anritsu MG3700A radio transmitter, operating from 50 Hz to 6 GHz, covers the frequency range required for the experiment (800 MHz, 1800 MHz, and 2600 MHz bands). This guarantees that the transmitter can generate signals within the specific bands necessary for the experiment.

The Pasternack antennas selected provide high directivity and gain, enhancing the efficiency of signal transmission. The directivity of these antennas ensures reliable data collection for the frequencies of interest.

The Anritsu MS2721B spectrum analyzer is capable of measuring signals in the frequency range of 9 KHz to 7.1 GHz, covering all bands used in the experiment. Additionally, the integrated GPS feature provides precise geographical location data for each measured point.

The transmitting antenna was installed at a height of 20 meters, while the receiving antenna was mounted on the roof of a vehicle, at a total height of 3.6 meters from the ground (Figure 1). The vehicle traveled along three different routes near the university (Figure 2), maintaining a constant speed of 20 km/h.

Figure 1 –Vehicle used in the measurement campaign, with the receiving antenna installed on its roof

Source: authors' files

Figure 2 –Map showing the three routes covered in the measurement campaign. Route 1 includes Cel. João Medeiros Street and Passeio dos Girassóis Avenue, while Routes 2 and 3 correspond to Cap. Mor Gouveia and Miguel Alcides de Araújo Avenues, respectively
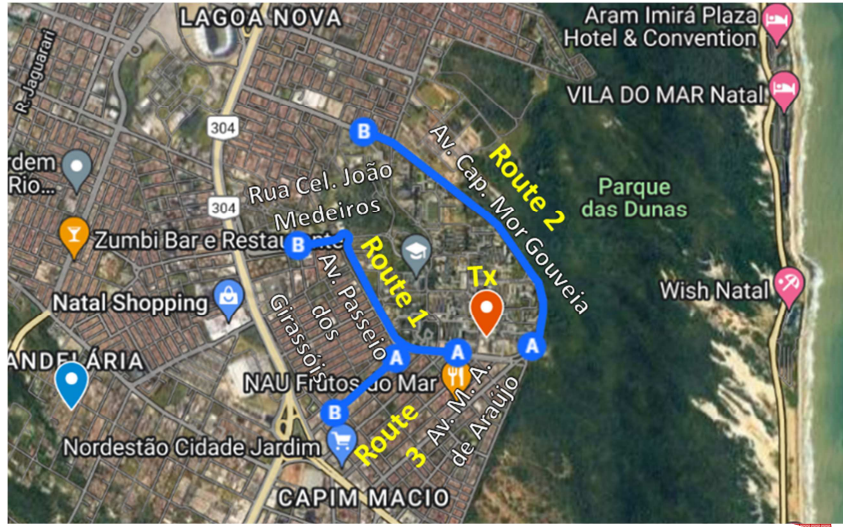
Source: Google Maps, 2024

After collecting the experimental data, it was necessary to convert the measured power into propagation loss. This conversion was performed using Equation (1):

$$PL = P_t + G_t + G_r - L_C + P_m \tag{1}$$

where PL represents the propagation loss, P_t is the transmitter's power output, G_t and G_r are the gain of the transmitting and receiving antennas, respectively, and L_C is the loss in the cables. Finally, P_m denotes the measured power.

An important factor that indicating the rate at which path loss decreases over distance is the path loss exponent, γ. This value usually typically ranges between 2 and 4 (Weinstock, 2007). A value of γ = 2 would correspond to a free space environment (straight line of sight, no obstructions), while γ = 4 represents an environment with high losses. Table 2 presents the losses values for each route and frequency.

Table 2 – Path loss exponents by route and frequency band

Route	Frequency band	γ
1	800 MHZ	2.4
	1800 MHz	2.5
	2600 MHz	2.9
2	800 MHz	2.4
	1800 MHz	3.4
	2600 MHz	3.1
3	800 MHz	2.3
	1800 MHz	2.7
	2600 MHz	2.8

Source: Weinstock, 2007.

The measured path loss exponent values align with those reported in the literature, confirming the intermediate characteristics of the suburban environment, which lies between free space and high-loss environments. Suburban areas typically present a moderate level of obstructions, as reflected in the measured γ values.

To ensure data accuracy and maintain the integrity of the dataset, each detected outlier was replaced with the previous non-outlier value. This approach provides a reasonable estimate of the expected signal strength at those points, preserving data continuity and trends while minimizing the introduction of erroneous values.

3 Numerical methods

Artificial Neural Networks (ANNs) are mathematical-computational models inspired by the neural structure of intelligent organisms and their capacity to acquire knowledge through experience. ANNs implement a nonlinear mapping of a set of inputs to a set of outputs, executed through layers of neurons. The input values are multiplied by specific synaptic weights of each layer, aiming to produce an appropriate output based on the inputs (Popescu; Nafornita; Constantinou, 2005). The approach using neural networks in this study utilizes two different methods:

- The first model uses two distinct inputs: one representing the distances between the transmitter and receiver, and another representing the error difference between experimental data and values predicted by the empirical propagation model (the Ericsson model was selected after performance tests). The output consists of the values obtained from the measurements. These values correspond to data collected across all routes.

  This method was chosen for its hybrid nature, which combines empirical propagation models with the learning capability of ANNs. This approach is particularly valuable in scenarios where traditional empirical models may not adequately capture the complexities of the environment;

- The second model employs four inputs, which are derived from terrain characteristics and the measurement setup: the distances between the transmitter and receiver, the azimuth angle between the transmitter and receiver, the altitude of the receiver, and the average terrain elevation. The output node represents the controlled measurement loss.

  This model was selected to explore the influence of terrain characteristics on signal propagation, similar to methodologies used in previous studies cited in the Introduction. By including these variables, the network can learn from relevant contextual data, improving the accuracy of predictions in environments where topography significantly impacts signal propagation.

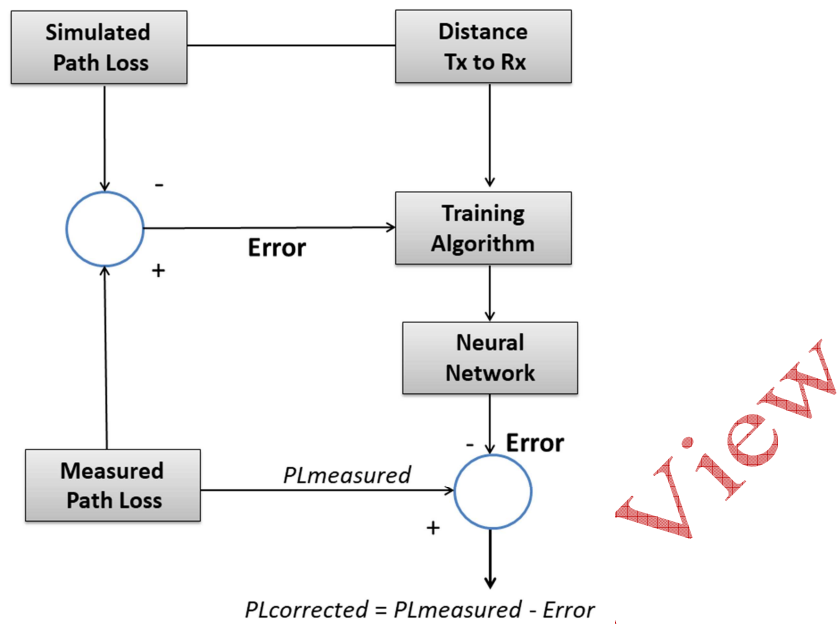*A. Error-Based Neural Network – EBNN*

The input to this model comprises two vectors: 1) the distance between the transmitter and receiver, and 2) the error between the measured values and those calculated by a selected propagation model:

$$Error = PL_{measured} - PL_{predicted} \qquad (2)$$

where $PL_{measured}$ represents the propagation loss obtained though field measurements, and $PL_{predicted}$ is the propagation loss calculated by Ericsson model.

The output vector, also known as the neural network target, consists of the measured path loss. The training phase of the neural network structure is shown in Figure 3, while the network architecture is illustrated in Figure 4.

Figure 3 – Error-Based Neural Network (EBNN) model training process

Figure 4 – Artificial Neural Network (ANN) applied to the Error-Based Neural Network (EBNN) model
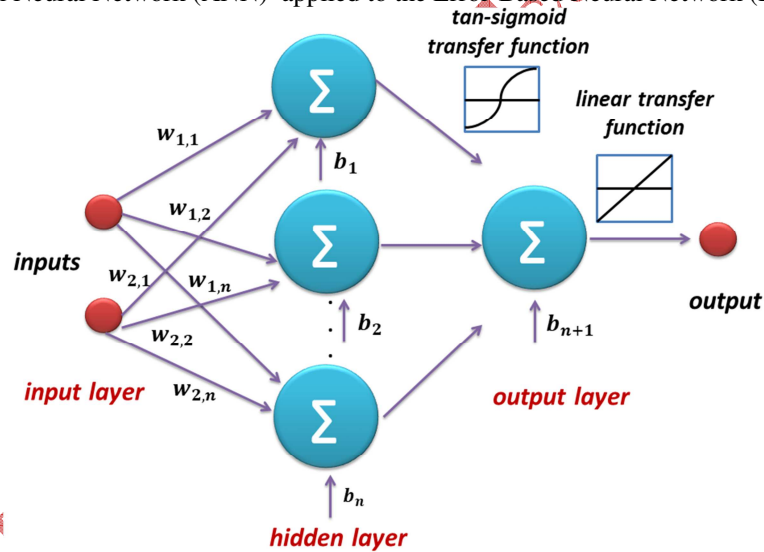
The designed ANN is a multilayer perceptron type, consisting of two inputs and one output, with a single hidden layer. The input set includes two vectors with values collected along three routes for each frequency: 560 elements for the 800 MHz scenario, 357 elements for the 1800 MHz, and 519 elements for the 2600 MHz frequency. The transfer functions used for the hidden and output layers are tangent sigmoid and linear, respectively, with the Levenberg-Marquardt algorithm chosen for network training.

In this network, each of the two inputs (the distance between the transmitter and receiver, and the error between measured and predicted values) is fed into the hidden layer, where neurons compute their weighted sum (w), add a bias (b), and apply a tangent sigmoid ($tanh$) activation function.

Synaptic weights are values associated with inputs, adjusted during training to improve the model results and store the acquired knowledge. The bias increases or decreases the input of the activation function depending on whether this value is negative or positive (Haykin, 2009). The activation function introduces non-linearity, enabling the network to capture complex patterns in the data.

The outputs from the hidden layer neurons are then passed to the output layer. This layer consists of a single neuron that calculates a weighted sum of the hidden layer outputs, adds another bias, and applies a linear activation function. The linear function ensures that the final output is a direct result of the weighted sum, suitable for regression tasks requiring precise numerical outputs.

Alternatives such as the standard backpropagation algorithm and the Levenberg-Marquardt method were considered, with the latter proving more efficient in terms of convergence speed (Haykin, 2009). The stopping criterion was set based on the number of epochs, capped at 1000, as convergence significantly decreased beyond this value.

To prevent overfitting and ensure the network's generalization capability, the data was divided into three sets: training (60%), validation (25%), and testing (15%). The training set was used to adjust network weights. The validation set assessed the network's generalization capacity, also serving as a stopping criterion using a cross-validation strategy. Finally, the testing set provided a realistic estimate of network performance on new data.

Network parameters were optimized using a cross-validation approach, where the dataset was partitioned into multiple subsets. In each iteration, the model was trained on all but one of these subsets, with the remaining subset used for testing. This process was repeated multiple times, rotating the test subset at each round. The final accuracy was determined by averaging the results obtained from all iterations (Dai, 2013).

To limit the number of models applied to the error-based neural network, a test methodology involving all propagation models was employed. Ultimately, the Ericsson model demonstrated the best average performance and was selected for use in the evaluation process.

B. Terrain parameters Based Neural Network – TBNN

This model shares the same architecture as depicted in Figure 4. The transfer functions for the hidden and output layers remain the tangent sigmoid and linear functions, respectively, and the Levenberg-Marquardt algorithm continues to be used for training. The primary difference lies in the inputs. The TBNN model incorporates four distinct inputs related to terrain and propagation setup (across the three routes):

- Distances between transmitter (Tx) and receiver (Rx);
- Receiver altitude;
- Average terrain elevation (origin and destination);
- Azimuth angle between Tx and Rx.

C. Propagation models

The path loss was calculated using the Ericsson and TR 36942 models, which covered all frequency bands, as well as the SUI and ECC-33 models, applied to the 1800 MHz and 2600 MHz bands. These propagation models were selected due to their relevance, applicability to the specific environment and frequency bands of interest, and their use in several propagation studies, such as those by Kumari $et\ al.$ (2011), Mathew, George, and Pereira (2017), and Siddiqui, Fatima, and Ahmad (2019).

Table 3 presents the main equations associated with these models (3GPP, 2024a; Mathew; George; Pereira, 2017; Kumari $et\ al.$, 2011).

Table 3 – Path loss models equations

Model	Equations	
3GPP TR 25.942	$PL_{TR} = 40[1 - 0.004h_{Tx}]log_{10}(d) - 18log_{10}(h_{Tx}) + 21log_{10}(f) + 80$	(3)
Ericsson	$PL_E = a_0 + a_1 log_{10}(d) + a_2 log_{10}h_{Rx} + a_3 log_{10}h_{Tx}log_{10}(d) - 3.2(log_{10}11.75h_{Rx})^2 + g(f)$	(4)
	$g(f) = 44.9log_{10}(f) - 4.78[log_{10}(f)]$	(5)
ECC-33	$PL_{ECC} = A_{fs} + A_{bm} - G_b - G_r$	(6)
	$A_{fs} = 92.4 + 20log_{10}(d) + 20log_{10}(f)$	(7)
	$A_{bm} = 20.41 + 9.83log_{10}(d) + 7.894log_{10}(f) + 9.56[log_{10}(f)]^2$	(8)

		$G_b = log_{10}\left(\dfrac{h_{Tx}}{200}\right)\{13.958 + 5.8[log_{10}(d)]^2\}$	(9)
		$G_r = [42.47 + 13.7log_{10}(f)log_{10}(f)][log_{10}(h_{RX}) - 0.585]$	(10)
		$G_r = 0.759(h_{Rx}) - 1.862$	(11)
SUI		$PL_{SUI} = A + 10\gamma log_{10}\left(\dfrac{d}{d_0}\right) + \Delta L_f + \Delta L_h + s$	(12)
		$A = 20log_{10}(\dfrac{4\pi d_0}{\lambda})$	(13)
		$\gamma = a - bh_b + \dfrac{c}{h_{Tx}}$	(14)
		$\Delta L_f = 6log_{10}(\dfrac{f}{2000})$	(15)
		$\Delta L_h = -10,8log_{10}\left(\dfrac{h_{Rx}}{2}\right)$	(16)
		$\Delta L_h = -20log_{10}(\dfrac{h_{Rx}}{2})$	(17)
		$s = 0,65[log_{10}(f)]^2 - 1,3 + log_{10}(f) + \alpha$	(18)

Source: 3GPP TR 25.942 (2024a); Mathew, George; Pereira (2017); Kumari *et al.* (2011)

where *d* represents the distance between the transmitter and receiver, *f* is the frequency in MHz, h_{Tx} is the height of the transmission antenna in meters, and h_{Rx} denotes the height of the reception antenna in meters. PL_{TR}, PL_E, PL_{ECC} and PL_{SUI} represent the path loss for TR 25.942, Ericsson, ECC-33 and SUI models, respectively.

Regarding the Ericsson model equations, $g(f)$ is the frequency correction factor, and a_0, a_1, a_2 and a_3 are constants that can be adjusted according to the scenario. The default values, which were used in the Ericsson model employed in this study, are 36.2, 30.2, 12.0, and 0.1, respectively.

In the ECC equations, A_{fs} represents the free space attenuation (in dB), while A_{bm} denotes the average path loss (in dB), which depends on the distance (in km) from the base station antenna to the user antenna and the operating frequency *f* (in GHz). Additionally, G_b is the gain (dB) of the base station antenna at a certain height (m), and G_r refers to the gain of the receiver antenna Equation 10 applies to small or medium cities, while Equation 11 applies to large cities.

For the SUI equations, *A* represents the free space path loss, ΔL_f is the frequency correction factor, and ΔL_h is the receiver antenna height correction factor. Equation 16 is valid for terrain types A (areas with the highest path loss, such as densely populated urban regions) and B (terrains with moderate path loss, typical of suburban environments), while Equation 17 is used for terrain type C (areas with the least path loss, such as rural or flat region). λ denotes the wavelength (m), and *s* represents the shadowing effect, where α equals 5.2 dB for urban and suburban environments (terrain types A and B) and 6.6 dB for rural environments (terrain C). Finally, γ is the propagation loss exponent for the different types of terrain.

4 Results and discuss

This section presents the performance analysis of the evaluated methods. RMSE serves as a metric that quantifies the discrepancy between the values predicted by the models and those measured. A lower RMSE is preferable, as it indicates that the model's predictions are closer to the real data, while higher values suggest that the model may not accurately capture the relationship between the data.

Statistically, the parameter *p* determines the level of significance in the experiment, with higher *p*-values indicating greater similarity between the data distribution and the experimental data. Any result with a p-value below 10^{-6} is considered null. The "Significance" column, based on the *p*-value, indicates whether the model is deemed accurate and statistically equivalent to real data. If the *p*-value is less than 0.05, the method is categorized as Significantly Different (S.D.); otherwise, it is classified as Significantly Equivalent (S.E.) (Andrade, 2019).

The *p*-value is used to assess the null hypothesis that the model's performance aligns with the expected performance based on measured data. A low *p*-value (less than 0.05) suggests that the differences between the model and real data are statistically significant and not due to random chance.

Conversely, a high *p*-value indicates that the model can be considered statistically equivalent to the real data, within the margin of statistical error.

A. 800 MHz scenario

The performance results for the 800 MHz band are highlighted in Table 4. Both neural network models were able to reduce the Root Mean Square Error (RMSE) while increasing the *p*-value, thereby minimizing the discrepancy between simulated and measured path loss data. In the case of the Error-Based Neural Network (EBNN), the RMSE approached zero. In terms of propagation models, both methods demonstrated regular performance, with TR 25.942 showing relatively better results. However, this regular performance was not confirmed in the similarity tests, where both conventional models exhibited low correlation between the measured and optimized values.

Table 4 – Performance of methods at 800 MHz

| Route	Model	RMSE(dB) 800 MHz	Wilcoxon rank-sum *p*-test 800 MHz	Significance @95% 800 MHz
1	Ericsson	10.76	null	S.D
	TR 25.942	8.67	0.001	S.D
	EBNN	0.003	0.984	S.E
	TBNN	5.082	0.675	S.E
2	Ericsson	11.18	null	S.D
	TR 25.942	7.63	0.001	S.D
	EBNN	0.626	0.997	S.E
	TBNN	5.96	0.541	S.E
3	Ericsson	10.817	null	S.D
	TR 25.942	9.44	null	S.D
	EBNN	0.003	0.972	S.E
	TBNN	5.7	0.555	S.E

Source: research data

These results are further supported by the analysis of Figures 5 and 6. Figure 5 presents the graph of the propagation loss as a function of distance, and Figure 6 provides a box-and-whisker plot comparing the datasets obtained by each method against the data collected from the measurement campaign, specifically for route 1.

Figure 5 –Metric comparison of path loss models and campaign measurements at 800 MHz for Route 1
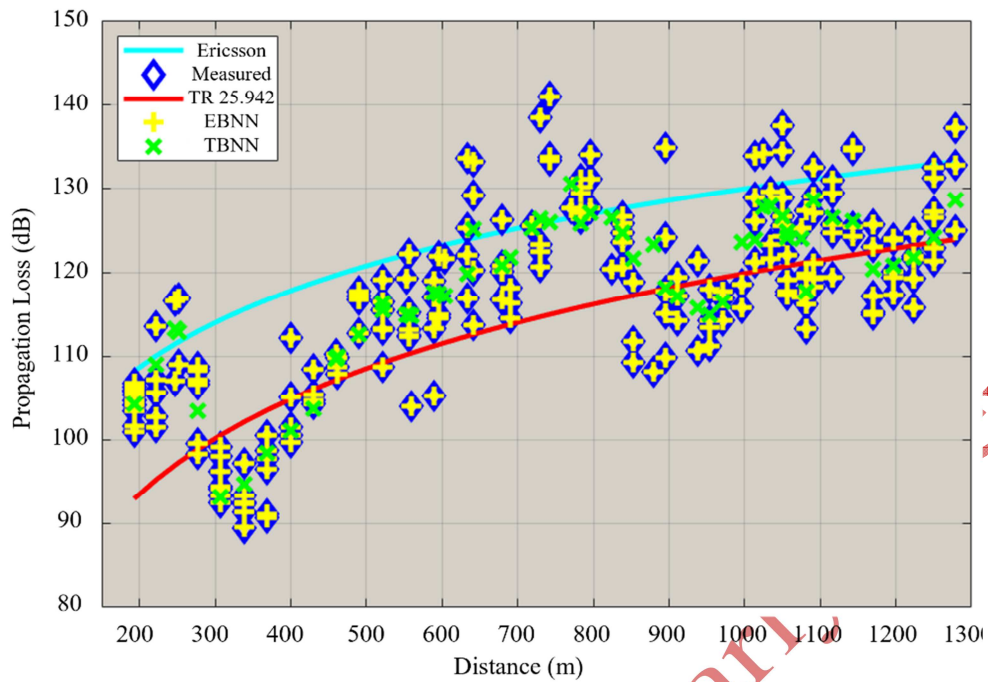
Source: research data

Figure 6 – Box-and-whisker plot comparing different methods for Route 1 (800 MHz)
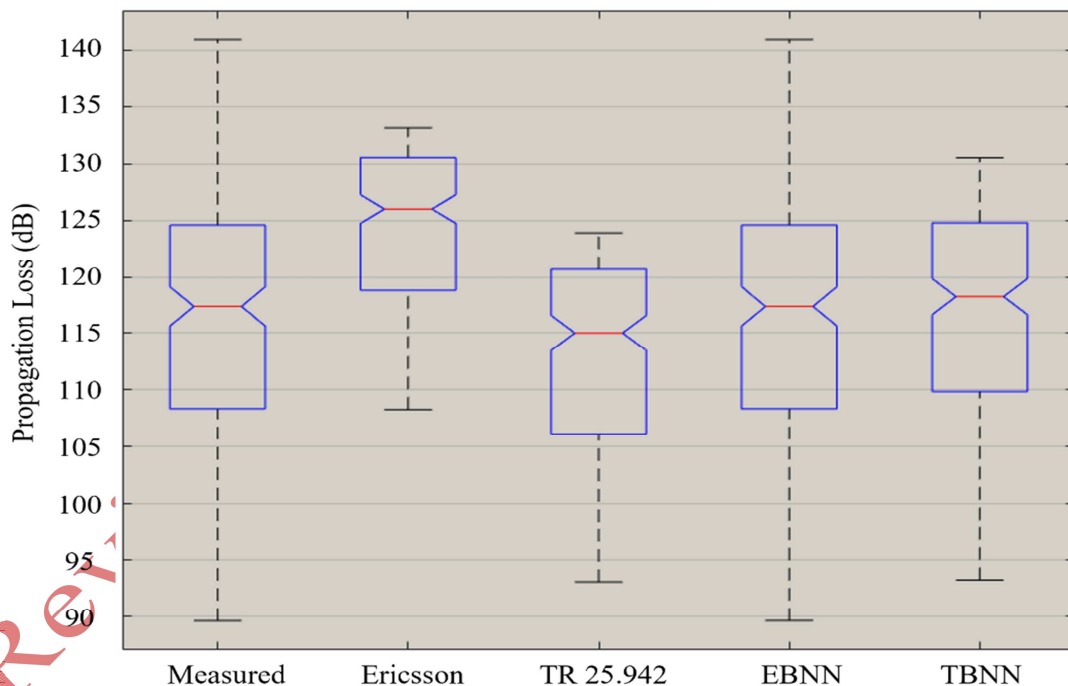
Source: research data

For the terrain-based network model in this scenario, while the distribution was slightly flatter, it closely aligned with the interquartile range, indicating high similarity with the measured data. However, this model demonstrated lower fidelity in data distribution, as its range occupied only part of the measured data sampling space (resulting in a *p*-value of 0.675). The EBNN model addressed these issues by achieving a data distribution closely matching the measured values, obtaining a *p*-value of 0.984.

*B. 1800 MHz and 2600 MHz scenario*

For measurements at 1800 MHz, ANN techniques again yielded the best performance across all metrics, as shown in Table 5. The EBNN approach achieved an RMSE close to zero and a *p*-value near 1, reinforcing its similarity to the collected data. Among the conventional models, the ECC model demonstrated the best performance across all routes, with the Ericsson model performing well specifically for route 2 (as illustrated in Figures 7 and 8).

Table 5 – Performance of analyzed methods at 1800 MHz

| Route	Model	RMSE(dB) 1800 MHz	Wilcoxon rank-sum *p*-test 1800 MHz	Significance @95% 1800 MHz
1	Ericsson	11.3	null	S.D
	TR 25.942	13.87	null	S.D
	SUI	11.89	null	S.D
	ECC	7.82	0.057	S.E
	EBNN	0.005	0.993	S.E
	TBNN	4.94	0.384	S.E
2	Ericsson	6.857	0.008	S.D
	TR 25.942	16.56	null	S.D
	SUI	13.727	null	S.D
	ECC	7.827	null	S.D
	EBNN	0.001	0.998	S.E
	TBNN	4.97	0.634	S.E
3	Ericsson	12.515	null	S.D
	TR 25.942	12.45	null	S.D
	SUI	9.58	null	S.D
	ECC	8.4	0.376	S.E
	EBNN	0.001	0.983	S.E
	TBNN	6.586	0.708	S.E

Source: research data

Figure 7 –Metric comparison of path loss models and campaign measurements at 1800 MHz for Route 2
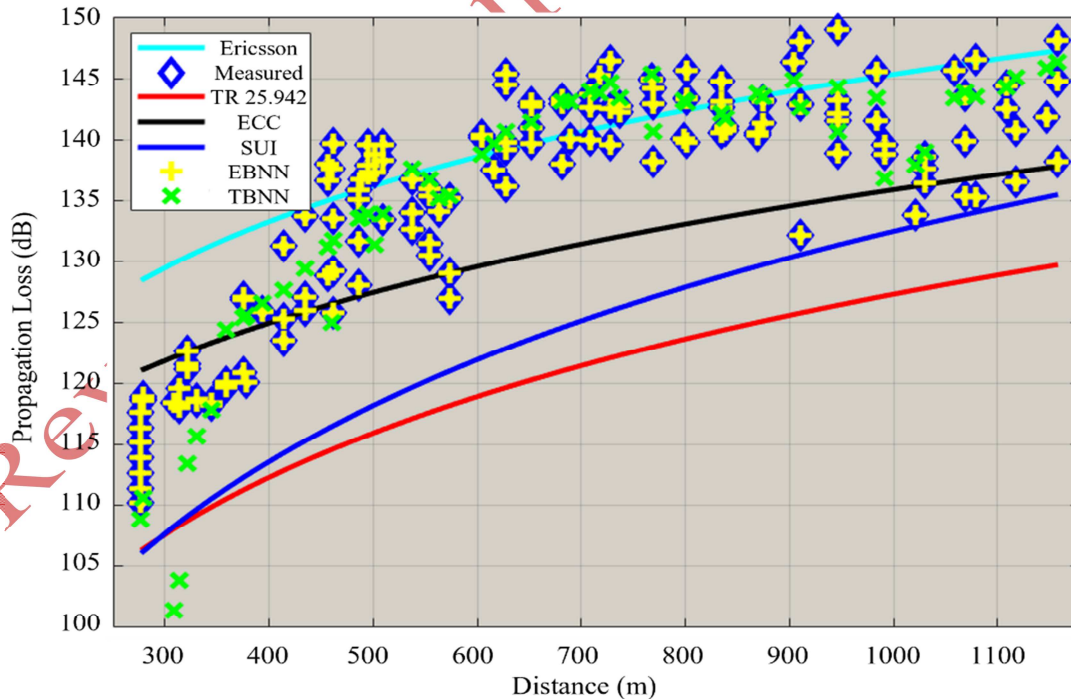
Source: research data

Figure 8 – Box-and-whisker plot comparing different methods for Route 2 (1800 MHz)
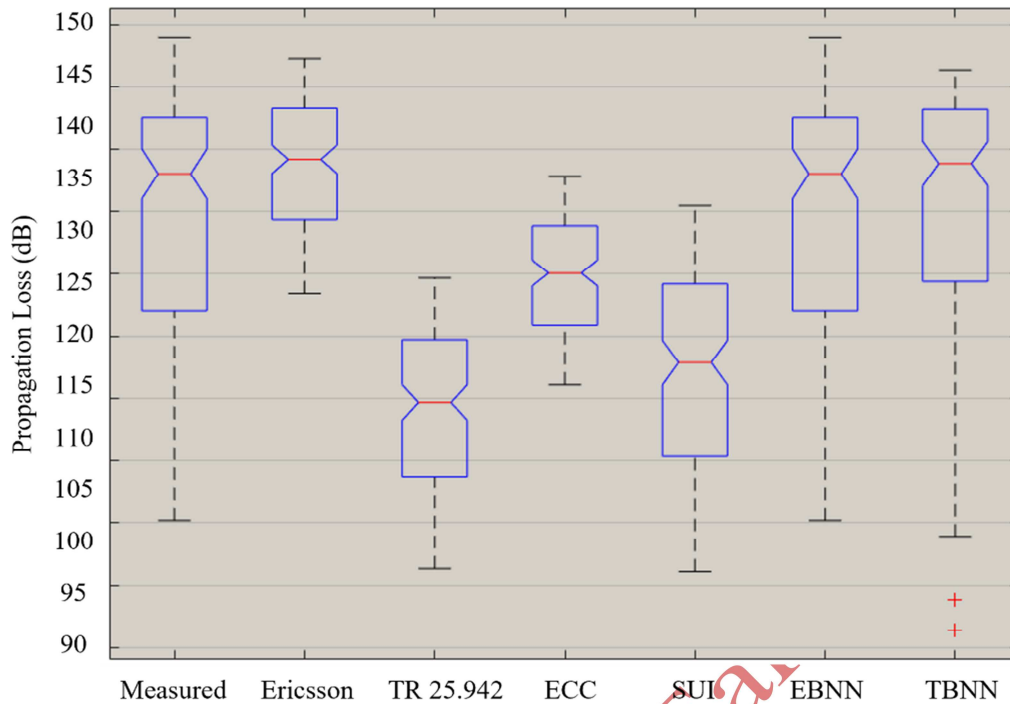
Conversely, the SUI and TR 25.942 models produced below-average results, deviating significantly from the measured data points along the entire path and yielding low RMSE values. The box-and-whisker plot representation reveals a greater dispersion when compared to experimental data.

In the 2600 MHz scenario for route 3, an increase in data dispersion is observed (Figures 9 and 10). This further highlights the importance of employing approaches capable of performing nonlinear mapping of the samples, as is the case with neural networks. Path loss models demonstrated low similarity (Table 6), underscoring the challenge faced by linear models in accurately handling data samples with significant dispersion. Even predictions using curve optimization techniques, such as genetic algorithms or the least squares method, would likely not resolve these discrepancies.

Table 6 – Performance of analyzed methods at 2600 MHz

Route	Model	RMSE(dB) 2600 MHz	Wilcoxon rank-sum p-test 2600 MHz	Significance @95% 2600 MHz
1	Ericsson	14.85	null	S.D
	TR 25.942	12.92	null	S.D
	SUI	9.99	null	S.D
	ECC	8.9	0.167	S.E
	EBNN	0.003	0.999	S.E
	TBNN	5.32	0.987	S.E
2	Ericsson	9.677	0.001	S.D
	TR 25.942	19.19	null	S.D
	SUI	15.01	null	S.D
	ECC	10.16	null	S.D
	EBNN	0.012	0.995	S.E
	TBNN	5.46	0.605	S.E
3	Ericsson	14.83	null	S.D
	TR 25.942	16.86	null	S.D
	SUI	13.77	null	S.D
	ECC	11.27	0.121	S.E
	EBNN	0.003	0.971	S.E
	TBNN	8.45	0.777	S.E

Figure 9 –Metric comparison of path loss models and campaign measurements at 2600 MHz for Route 3
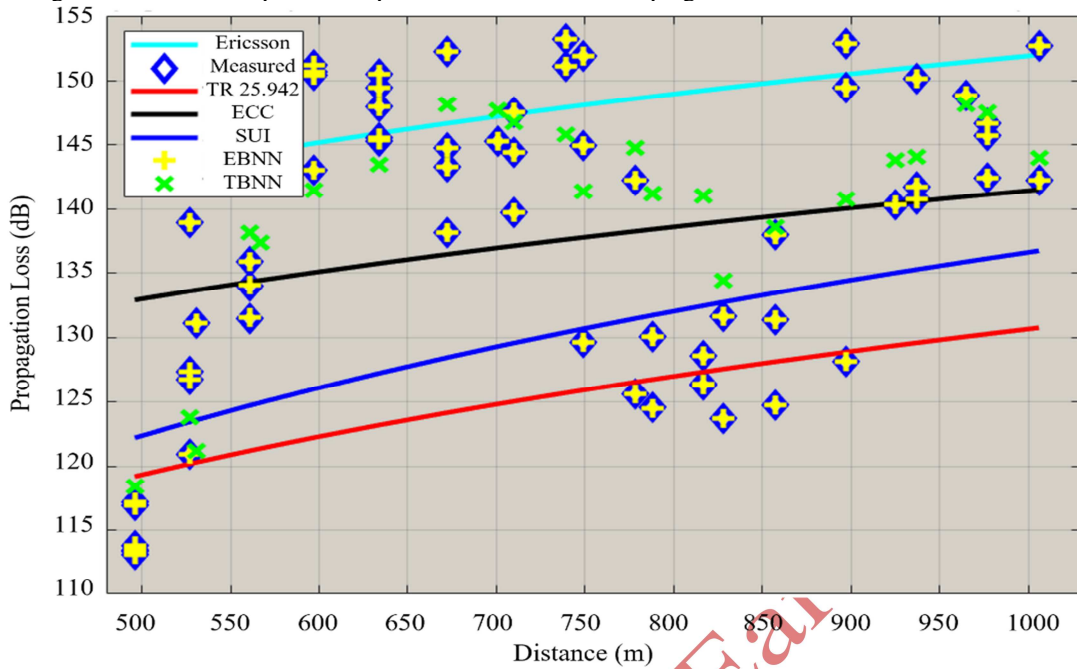
Figure 10 – Box-and-whisker plot comparing different methods for Route 3 (2600 MHz)
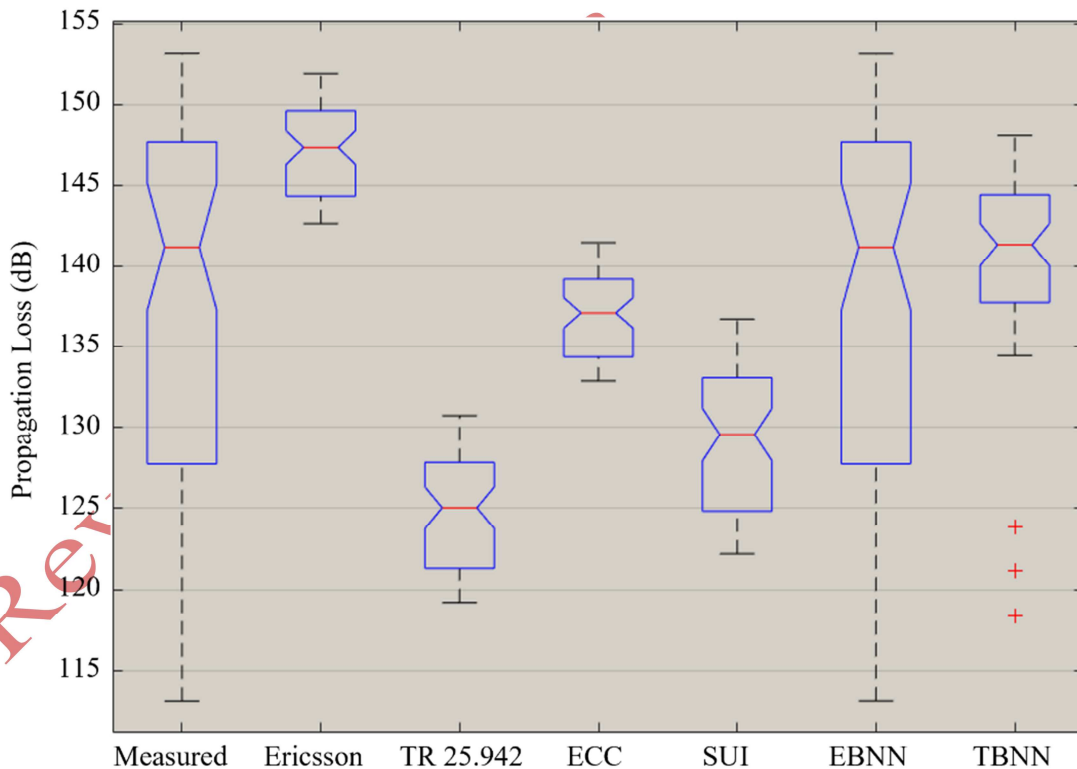
5 Conclusions

This study analyzed two types of artificial neural network (ANN) approaches for predicting propagation loss at 800 MHz, 1.8 GHz, and 2.6 GHz in suburban areas: an error-based hybrid model that utilizes the error calculated by the Ericsson model, and another neural network that incorporates

terrain features as input parameters. The performance of these techniques was evaluated and compared using Root Mean Square Error (RMSE) and their similarity with well-established path loss models from the literature. The following conclusions can be drawn from the results:

- Both ANN approaches successfully reduced RMSE values while increasing p-values, significantly decreasing the difference between simulated and measured values. In the case of the Error-Based Neural Network (EBNN), RMSE values approached zero. This demonstrates the neural networks' capacity to learn and incorporate non-deterministic values present in experimental data;
- In most cases, traditional path loss modeling methods resulted in *p*-values less than 0.05. This is likely due to the greater dispersion of measured data, particularly on routes 2 and 3, even after the replacement of outliers;
- Based on the results obtained from classical models, it is recommended to use the TR 25.942 model for the 800 MHz band, the ECC model for the 1800 MHz band, and both the SUI and ECC models for the 2600 MHz band;
- The most efficient technique, characterized by RMSE values close to zero and p-values near 1, recommended for predicting propagation loss in the scenarios analyzed, was the error-based ANN;
- Compared to the studies by Popescu, Nafornita, and Constantinou (2005) and Sanches and Cavalcante (2015), this research covered a broader range of frequencies and achieved lower average RMSE values. The results are also notable when compared to more recent studies, such as those by Jo *et al.* (2020) and Popoola *et al.* (2021);
- Both neural networks proved to be effective and accurate prediction methods, demonstrating simulation results closely aligned with actual field measurements. These methods are valuable tools for LTE and 5G network designers, providing reliable predictions of propagation loss;
- The high accuracy of the predictions allows for a more precise design of network coverage, minimizing areas with poor signal reception and avoiding excessive cell overlap. This level of precision enables the optimal allocation of resources, including the strategic positioning of base stations and the adjustment of transmission power, resulting in significant cost savings and improved network performance.

A noteworthy aspect of this study is the use of the Wilcoxon rank sum method, as opposed to the standard deviation, a tool frequently employed in propagation studies, such as those by Popescu, Nafornita, and Constantinou (2005), and Sanches and Cavalcante (2015). While the standard deviation measures the variability of prediction errors around the mean, providing a general assessment of the accuracy of the propagation loss model, the Wilcoxon test offers a more detailed and robust analysis of the similarity between the distributions of predicted and measured errors. The standard deviation can be influenced by outliers and non-normal distributions (Riedel *et al.*, 2019), potentially skewing the evaluation of model accuracy. In contrast, the Wilcoxon test is less sensitive to these factors, as it is based on ranks rather than raw values.

It is also important to note that the neural network models (EBNN and TBNN) were specifically trained for suburban environments and the frequency bands considered in this study. As a result, their performance may not be as effective in urban or rural settings.

Future research directions are clear. The proposed method could be further evaluated using different datasets, including those from urban and rural areas, and for various frequency bands, including 5G systems. Additionally, it can be extended to future 6G networks, which are expected to integrate satellite systems, aerial networks, terrestrial communications, maritime communications, and underwater systems. In this context, Artificial Intelligence and Machine Learning techniques should be utilized to further enhance path loss prediction metrics.

References
3GPP. **TR 25.942**. 3rd generation partnership project. Technical specification group radio access network. Radio Frequency (RF) system scenarios. V18.0.0, March. 2024a. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=1362 . Acessed on: 06 sep. 2024.

3GPP. **TS 36.101.** 3rd generation partnership project. Technical specification group radio access network. Evolved universal terrestrial radio access (E-UTRA). User equipment (UE) radio transmission and reception. V18.6.0, July, 2024b. Available at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2411 . Acessed on: 06 sep. 2024.

3GPP. **TR 36.942**. 3rd generation partnership project. Technical specification group radio access network. Evolved universal terrestrial radio access (E-UTRA). Radio Frequency (RF) system scenarios. V18.0.0, April. 2024c. Avaliable at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=2592 . Acessed on: 06 sep. 2024.

3GPP. **TS 38.101-1**. 3rd generation partnership project. Technical specification group radio access network. NR. User Equipment (UE) radio transmission and reception. Part 1: range standalone. V18.6.0, July, 2024d. Avaliable at: https://portal.3gpp.org/desktopmodules/Specifications/SpecificationDetails.aspx?specificationId=3283 . Acessed on: 06 sep. 2024.

ANDRADE, C. The P value and statistical significance: misunderstandings, explanations, challenges, and alternatives. **Indian Journal of Psychological Medicine**, v. 41, n. 3, p. 210-215, 2019. DOI: https://doi.org/10.4103/ijpsym.ijpsym_193_19.

ASHRAFIJOO, B.; FARAHMAND, N. F-H.; MATIN, Y. A.; RAHMANI, K. Designing an optimal model using artificial neural networks to predict non-linear time series (case study: Tehran Stock Exchange Index). **Journal of System Management**, v. 8, n. 4, p. 65-80, 2022. DOI: https://doi.org/10.30495/jsm.2022.1965914.1679.

BENMUS, T. A.; ABBOUD, R.; SHATTER, M. K. Neural network approach to model the propagation path loss for great Tripoli area at 900, 1800, and 2100 MHz bands. In: 2015 INTERNATIONAL CONFERENCE ON SCIENCES AND TECHNIQUES OF AUTOMATIC CONTROL AND COMPUTER ENGINEERING (STA), 16., 2015, Monastir. **Proceedings [...]** Monastir: IEEE, 2015. p. 793-798. DOI: https://doi.org/10.1109/STA.2015.7505236.

CINTRA, R. S.; VELHO, H. F. C. Global data assimilation using artificial neural networks in Speedy model. In: INTERNATIONAL SYMPOSIUM UNCERTAINTY QUANTIFICATION AND STOCHASTIC MODELING, 2012, São Sebastião. **Proceedings [...].** São Sebastião, 2012, p. 648-654. Available at: http://mtc-m16d.sid.inpe.br/col/sid.inpe.br/mtc-m19/2012/10.17.17.02/doc/106RCintra.pdf. Accessed on: 25 aug. 2024.

DAI, Q. A competitive ensemble pruning approach based on cross-validation technique. **Knowledge-Based Systems**, v. 37, p. 394-414, 2013. DOI: https://doi.org/10.1016/j.knosys.2012.08.024.

EICHIE, J. O.; OYEDUM, O. D..; AJEWOLE, M.; AIBINU, A. M. Comparative analysis of basic models and artificial neural network based model for path loss prediction. **Progress In Electromagnetics Research**, v. 61, p. 133-146, 2017. DOI: http://dx.doi.org/10.2528/PIERM17060601.

Google Maps. Available at: https://maps.app.goo.gl/dgGymVNzCGGyxRhW8. Accessed on: 06 sep. 2024.

HAYKIN, S. **Neural networks and learning machines**. Prentice Hall, 2009.

HERNÁNDEZ, G.; ZAMORA, E.; SOSSA, H.; TÉLLEZ, G.; FURLÁN, F. Hybrid neural networks for big data classification. **Neurocomputing**, v. 390, p. 327-340, 2020. DOI: https://doi.org/10.1016/j.neucom.2019.08.095.

JI, X.; HE, X.; LV, C.; LIU, Y.; WU, J. Adaptive-neural-network-based robust lateral motion control for autonomous vehicle at driving limits. **Control Engineering Practice**, v. 76, p. 41-53, 2018. DOI: https://doi.org/10.1016/j.conengprac.2018.04.007.

JO, H.-S.; PARK, C.; LEE, E.; CHOI, H. K.; PARK, J. Path loss prediction based on machine learning techniques: principal component analysis, artificial neural network, and Gaussian process. **Sensors**, v. 20, n. 7, 1927, 2020. DOI: https://doi.org/10.3390/s20071927.

KUMARI, M.; YADAV, T.; YADAV, P.; SHARMA, P. K.; SHARMA, D. Comparative study of path loss models in different environments. International Journal of Engineering Science and Technology, v. 3, n. 4, p. 2945-2949, 2011.

MA, S.; ZHANG, X.; JIA, C.; ZHAO, Z.; WANG, S.; WANG, S. Image and video compression with neural networks: a review. **IEEE Transactions on Circuits and Systems for Video Technology**, v. 30, n. 6, p. 1683-1698, 2019. DOI: https://doi.org/10.1109/TCSVT.2019.2910119.

MATHEW, B.; GEORGE, W.; PEREIRA, M. QoS Enhancement in 4G heterogeneous networks using Kalman Filter & EWMA. **International Journal of Electronics and Communications Engineering and Technology (IJECET)**, v. 8, n. 3, p. 28-43, 2017. Available at: https://iaeme.com/Home/article_id/IJECET_08_03_004. Accessed on: 25 aug. 2024.

MORAITIS, N.; TSIPI, L.; VOUYIOUKAS, D. Machine learning-based methods for path loss prediction in urban environment for LTE networks. In: 2020 INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS (WiMob), 16., 2020, Thessaloniki. **Proceedings […].** Thessaloniki: IEEE, 2020. DOI: https://doi.org/10.1109/WiMob50308.2020.9253369.

PASTERNACK. **PE51043 and PE51054 antennas**. Pasternack. Available at: https://www.pasternack.com/panel-antenna-n-female-2500-2700-mhz-14-dbi-pe51043-p.aspx and https://www.everythingrf.com/products/all-antennas/pasternack-enterprises-inc/741-20-pe51054. Accessed on: 11 Aug. 2024.

POPESCU, I.; KANSTAS, A.; ANGELOU, E.; NAFORNITA, L.; CONSTANTIOU, P. Applications of generalized RBF-NN for path loss prediction. In: IEEE INTERNATIONAL SYMPOSIUM ON PERSONAL, INDOOR AND MOBILE RADIO COMMUNICATIONS, 13., Lisbon, 2002.

Proceedings […]. Lisbon: IEEE, p. 484-488, 2002. DOI: https://doi.org/10.1109/PIMRC.2002.1046748.

POPESCU, I.; NAFORNITA, I.; CONSTANTINOU, P. Comparison of neural network models for path loss prediction. *In*: IEEE INTERNATIONAL CONFERENCE ON WIRELESS AND MOBILE COMPUTING, NETWORKING AND COMMUNICATIONS (WiMob'2005), 2005, Montreal. **Proceedings […]**. Montreal: IEEE, 2005. p. 44-49. DOI: https://doi.org/10.1109/WIMOB.2005.1512814.

POPOOLA, S. I.; FARUK, N.; SURAJUDEEN-BAKINDE, N. T.; ATAYERO, A. A.; MISRA, S. Artificial neural network model for path loss predictions in the VHF band. In: TRIPATHI, M., UPADHYAYA, S. (eds) **Conference Proceedings of ICDLAIR2019**. ICDLAIR 2019. Lecture Notes in Networks and Systems, v. 175. Cham: Springer, 2019. DOI: https://doi.org/10.1007/978-3-030-67187-7_18.

RAJENDRA, P.; MURTHY, K. V. N.; SUBBARAO, A.; BOADH, R. Use of ANN models in the prediction of meteorological data. **Modeling Earth Systems and Environment**, v. 5, p. 1051-1058, 2019. DOI: https://doi.org/10.1007/s40808-019-00590-2.

RIEDEL, T.; TAKAHASHI, M.; TATSUKAWA, T.; ITOH, E. Evaluating applied flight-deck interval management using Monte Carlo simulations on the K-Supercomputer. **Transactions of the Japan Society for Aeronautical and Space Sciences**, v. 62, n. 6, p. 299-309, 2019. DOI: https://doi.org/10.2322/tjsass.62.299.

SAEED, M. A.; KHAN, M. Z.; KHAN, A.; SAEED, M. U.; HASSAN, M. A. S.; JAVED, T. Impact of propagation path loss by varying BTS height and frequency for combining multiple path loss approaches in macro-femto environment. **Arabian Journal for Science and Engineering**, v. 47, n. 2, p. 1227-1238, 2022. DOI: https://doi.org/10.1007/s13369-021-05819-w.

SANCHES, M. A. R.; CAVALCANTE, G. P. S. Modelos neuro-adaptados para prediçao de radiopropagaçao em sistemas móveis terrestres. **Revista da Sociedade Brasileira de Telecomunicações**, v. 1, n. 1, p. 11-15, 2015. Available at: https://jcis.sbrt.org.br/jcis/article/view/246. Accessed on: 25 aug. 2024. In Portuguese.

SIDDIQUI, S. A.; FATIMA, N.; AHMAD, A. Comparative analysis of propagation path loss models in LTE networks. In: 2019 INTERNATIONAL CONFERENCE ON POWER ELECTRONICS, CONTROL AND AUTOMATION (ICPECA), 2019, New Delhi. **Proceedings […].** New Delhi: IEEE, 2019. DOI: https://doi.org/10.1109/ICPECA47973.2019.8975464.

SILVA, P. H. F.; PASSOS, M. G. Numerical analysis of the SIGAnatel tool for technical feasibility studies of TV and FM broadcast channels. In: 2007 SBMO/IEEE MTT-S INTERNATIONAL MICROWAVE AND OPTOELECTRONICS CONFERENCE, 2007, Salvador. **Proceedings […].** Salvador: IEEE, 2007. p. 569-573.DOI: https://doi.org/10.1109/IMOC.2007.4404329.

SUN, S.; RAPPAPORT, T. S.; THOMAS, T. A.; GHOSH, A.; NGUYEN, H. C.; KOVÁCS, I. Z.; RODRIGUEUZ, I.; KOYMEN, O.; PARTYKA, A. Investigation of prediction accuracy, sensitivity, and parameter stability of large-scale propagation path loss models for 5G wireless communications. **IEEE Transactions on Vehicular Technology**, v. 65, n. 5, p. 2843-2860, 2016. DOI: https://doi.org/10.1109/TVT.2016.2543139.

SUNG, S.; CHOI, W.; KIM, H.; JUNG, J.-I. Deep learning-based path loss prediction for fifth-generation new radio vehicle communications. **IEEE Access**, v. 11, p. 75295-75310, 2023. DOI: https://doi.org/10.1109/ACCESS.2023.3297215.

WEINSTOCK, H. C. (ed.) **Focus on cognitive radio technology**. Nova Publishers, 2007.