Eksakta : Berkala Ilmiah Bidang MIPA
VOLUME 27 NO 03 2026, pp 298-315
ISSN : Print 1411-3724 — Online 2549-7464
DOI : https://doi.org/10.24036/eksakta/vol27-iss03/683

Eksakta
Berkala Ilmiah Bidang MIPA
http://www.eksakta.ppj.unp.ac.id/index.php/eksakta

Article

Residual-Based Analysis of the Mismatch Between LoRa Channel Models and Field Measurements in Urban and Rural Environments

Article Info

Article history :
Received April 25, 2026
Revised May 25, 2026
Accepted June 01, 2026
Published June 30, 2026

Keywords :
LoRa, path loss, propagation models, residual analysis, machine learning

Muhamad Bagus Fikril Alan[1], Pradini Puspitaningayu[1], Nurhayati Nurhayati[1*], Muhammad 'Aamir Nashrullah[1], Nobuo Funabiki[2], Erwin Sutanto[3], Fahmi Fahmi[4]

[1]Department of Electrical Engineering, Faculty of Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia
[2]Department of Electrical and Communication Engineering, Faculty of Engineering, Okayama University, Okayama, Japan
[3]Department of Physics, Faculty of Science and Technology, Universitas Airlangga, Surabaya, Indonesia
[4]Department of Electrical Engineering, Faculty of Engineering, Universitas Sumatera Utara, Medan, Indonesia

Abstract. Ideally, LoRa radio planning requires propagation models that can accurately estimate path loss across different environments using physically meaningful parameters. In real deployments, however, urban buildings, vegetation, terrain variation, and local obstruction often cause classical models to leave systematic errors that are not fully explained by distance alone. This study proposes a residual-based analysis to evaluate whether the mismatch between classical LoRa propagation models and field measurements contains deterministic patterns. The objective is to compare urban and rural LoRa measurements, optimize classical models, and examine residual structures using machine learning. Field measurements were conducted at an urban campus area of Universitas Negeri Surabaya and a rural area in Kediri, Indonesia. A total of 80 line-of-sight samples were collected from 10 m to 400 m using a 915 MHz LoRa system. Okumura-Hata, Close-In, and Floating-Intercept models were evaluated, and residuals from the optimized Close-In model were predicted using logarithmic distance, environment type, and obstacle density. The optimized FI model produced the lowest baseline RMSE, while autocorrelation confirmed structured residual behavior. Random Forest achieved the best residual prediction with $R^2 = 0.972$ and RMSE = 2.704 dB. These findings show that residual-based learning improves LoRa propagation interpretation for heterogeneous IoT planning tasks.

Corresponding Author :
Nurhayati Nurhayati
Department of Electrical Engineering, Universitas Negeri Surabaya, Surabaya, Indonesia
Email : nurhayati@unesa.ac.id

1. Introduction

The current Internet of Things (IoT) era continues to evolve rapidly, with application categories and security technologies receiving the highest publication attention compared to other networking aspects [1]. According to industrial statistical reports, the IoT market is projected to surpass $456 billion by 2025, with massive adoption expected in agriculture, healthcare, and smart cities by 2030 [2-4]. In this ecosystem, Low Power Wide Area Network (LPWAN) technologies serve as a fundamental pillar for providing connectivity in regions with minimal cellular infrastructure, particularly for long-range applications such as environmental monitoring and smart agriculture [5]. However, practical IoT implementation remains hindered by the lack of accurate radio planning guidelines, given that this technology is highly sensitive to environmental propagation conditions [6-8].

Long Range (LoRa) has emerged as a key LPWAN technology due to its advantages in transmission range, power efficiency, and low deployment costs [9-10]. Operating within the ISM frequency bands (433, 866, and 915 MHz) using Chirp Spread Spectrum (CSS) modulation, LoRa offers flexibility through Spreading Factor (SF) configurations that balance coverage and throughput [11-13]. Consequently, LoRa signal performance is significantly influenced by morphological factors such as buildings, vegetation, and topography [14]. Although literature evaluating LoRa propagation in both urban and rural environments is available, most studies remain fragmented; they typically focus on a single environment type without providing a comprehensive, head-to-head comparison between structured urban areas and open rural areas [15-18].

Recent studies show that LoRa propagation remains strongly dependent on environmental morphology. Azevedo and Mendonça reviewed LoRa propagation models and emphasized that accurate propagation modeling is essential for network planning across outdoor, indoor, and vegetation environments [19]. Lima et al. developed LoRaWAN propagation models for 915 MHz deployment in the Amazon region and showed that local morphology strongly affects RSSI and path loss behavior [20]. Mahjoub et al. also reported that vegetation density and site characteristics influence LoRa signal attenuation in agricultural environments [21].

However, recent studies still mainly focus on direct path loss modeling or empirical model fitting. Machine learning has increasingly been used in radio propagation modeling because it can learn nonlinear interactions between terrain, obstruction, and distance-related features [22]. Nevertheless, most LoRa studies have not specifically investigated whether the residual error left by classical propagation models contains deterministic patterns. This creates a research gap between model comparison and residual interpretation.

Therefore, the novelty of this study lies in analyzing LoRa propagation mismatch through a residual-based framework. This study does not use machine learning merely as a direct path loss predictor. Instead, machine learning is used as a diagnostic tool to examine whether residuals from physically grounded models can be explained by logarithmic distance, environment type, and obstacle density.

A fundamental weakness of classical propagation models such as Okumura-Hata, Close-in (CI), and Floating-intercept (FI) is their limitation in capturing heterogeneous environmental dynamics. The Okumura-Hata model often fails in areas with high vegetation density, while CI and FI models

frequently exhibit estimation bias or lose physical relevance to Free Space Path Loss (FSPL) at certain distances [23-25]. Although several studies have evaluated LoRa propagation in urban, suburban, rural, and vegetation-dominated environments, most of them primarily report prediction errors such as RMSE or mean absolute error to compare classical models. Such approaches implicitly treat the discrepancy between measured and predicted path loss as random noise. However, in heterogeneous environments, residual errors may arise from deterministic physical mechanisms, including multipath propagation, diffraction around obstacles, vegetation-induced attenuation, and shadowing caused by buildings or terrain morphology. Therefore, merely reporting aggregate error metrics is insufficient to explain why classical models fail under specific environmental conditions.

To address this gap, this study proposes a residual-based analysis framework for LoRa propagation modeling. Instead of using machine learning solely as a direct path loss predictor, this work employs machine learning as a diagnostic tool to examine whether residuals from classical propagation models contain learnable deterministic structures. By comparing urban and rural measurements under line-of-sight conditions and incorporating obstacle density as an environmental feature, this study provides a more detailed interpretation of model mismatch beyond conventional RMSE-based evaluation.

This article is organized as follows section II explains the research methodology, covering field data collection scenarios in urban and rural areas as well as the LoRa configuration parameters used. Section III presents the comparative analysis of RSSI and path loss measurements against theoretical models, discusses the residual analysis, and details the Machine Learning evaluation used to verify systematic patterns in model errors. Finally, Section IV summarizes the key findings and provides recommendations for future research development.

2. Method

2.1. Research Workflow

This research was conducted using a systematic approach designed to investigate the discrepancies between theoretical estimations and field realities. The research workflow is divided into key phases: field data acquisition, baseline propagation modeling, residual extraction and analysis, and pattern validation using Machine Learning. Visually, these research stages are illustrated in Figure 1.

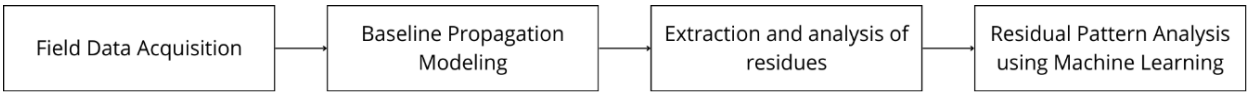

Figure 1. Research block diagram

2.2. Experimental Setup and Data Acquisition

Figure 2 illustrates the schematic diagram employed in this study, utilizing the Arduino Uno R4 WiFi microcontroller. This model represents the latest iteration of the Arduino series, built upon the Renesas RA4M1 architecture. The board is integrated with an Espressif ESP32-S3 WiFi module, offering a flash memory capacity of 256 KB [26].

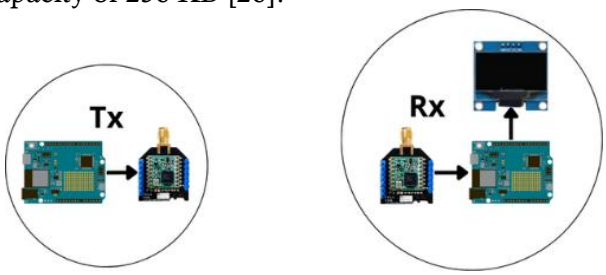

Figure 2. Hardware schematic diagram

For long-range communication, the system utilizes a Cosmic LoRa module equipped with the RFM95 IC, operating within the sub-GHz spectrum. LoRa employs the Chirp Spread Spectrum (CSS) modulation technique, which enables extended transmission ranges with low power consumption and robust immunity to interference [27-30]. Technical parameters such as Spreading Factor (SF) and Bandwidth play a pivotal role in determining system performance; the specific configurations used are detailed in Table 1 as follows:

Table 1. LoRa communication parameters used in the experiment

Parameters	Value
Spreading factor	7
Bandwidth	125 kHz
Transmission power	20 dBm
Frequency	915 MHz

Subsequently, field experiments were conducted in two distinct environments representing real-world conditions: urban and rural scenarios. The urban environment is characterized by a structured layout with high building density. The urban measurements were carried out at the State University of Surabaya campus, a location featuring significant physical obstructions, including multi-story buildings and vegetation. Conversely, the rural experiments took place in the rural area of Kediri, which exhibits an open terrain contour with minimal obstructions. This approach facilitates the evaluation of LoRa propagation performance under two extreme conditions that contrast significantly in terms of topography and structural composition.

2.3. Data Processing

The data acquisition workflow employed in this study is depicted in Figure 3, utilizing two Arduino Uno R4 units configured as the transmitter and receiver, respectively. The system initialization process encompasses the setup of the LoRa module, the NTPClient for precise time synchronization, an OLED LCD for real-time data visualization, and the Blynk platform connection. On the transmitter side, the system retrieves critical parameters, including the Spreading Factor (SF), bandwidth, the binary control status ('Yes' or 'No') from the Blynk interface, and the timestamp.

The system verifies whether the control status received from the Blynk interface corresponds to the 'Yes' command. Upon confirmation, the system initiates a dedicated data logging procedure to record the maximum and minimum RSSI values over a one-minute interval. This data collection methodology is designed to validate the accuracy of the path loss exponent parameters utilized in RSSI estimation models [31-32]. Subsequently, the captured data is displayed on the receiver unit, concurrently with latency, throughput metrics, and the associated timestamp.

The final dataset consisted of 80 measurement samples, comprising 40 samples from the urban environment and 40 samples from the rural environment. Measurements were conducted from 10 m to 400 m with a 10 m interval under line-of-sight conditions. Since all measurement points were acquired under LoS conditions, the LoS status was treated as a controlled measurement condition rather than a varying input feature. In addition to distance and environment type, obstacle density was introduced as an additional environmental feature and categorized into three levels: low, medium, and high. The distance variable was transformed into logarithmic form using $\log_{10}(d)$, consistent with the log-distance behavior of radio propagation.

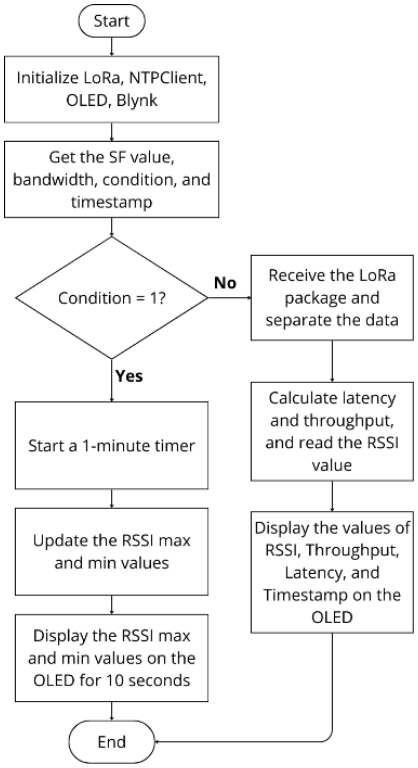

Figure 3. Data collection flowchart

The environment variable was encoded as a binary feature, where 0 represents the rural environment and 1 represents the urban environment. Obstacle density was ordinally encoded as 0, 1, and 2 for low, medium, and high density, respectively. The resulting input features for machine learning were logarithmic distance, environment type, and obstacle density. Due to the limited dataset size, model evaluation was not based solely on a single hold-out split. A 5-fold cross-validation strategy was applied to reduce evaluation bias and provide a more robust estimate of model generalization performance. In each fold, the dataset was divided into training and validation subsets, and the average R^2 and RMSE values were reported.

A limitation of this dataset is the absence of GPS-based coordinate information and explicit NLoS measurement samples. Therefore, the present analysis focuses on distance-based, environment-based, and obstacle-density-based residual characterization. Future work should incorporate spatial coordinates and both LoS/NLoS classifications to enable more detailed spatial residual modeling.

2.4. Propagation Modeling

This study employs both deterministic and empirical models as baselines to evaluate the propagation characteristics of the LoRa channel and to quantify path loss residuals. The models selected for analysis include Okumura-Hata, Close-In (CI), and Floating-Intercept (FI) [16], [25]. In addition to utilizing standard parameters, this research conducts parameter optimization (tuning) on selected models based on empirical measurement data. This optimization employs linear regression techniques specifically to minimize the Mean Square Error (MSE). These models were selected due to their representativeness of diverse environmental conditions. The Okumura-Hata model is utilized to predict signal attenuation in both urban and rural environments for frequencies below 1500 MHz. The governing equation is expressed as follows [16]:

$$Lp|dB = 69.55 + 26.16\,log_{10}\,f - 13.82\,log_{10}\,ht - a(hr) + (44.9 - 6.55\,log_{10}\,ht)\,log_{10}\,d \quad (1)$$

$$a(hr) = 3.2\,[log_{10}(\,11.75 * hr]^2 - 4.97 \quad (2)$$

where,
Lp(dB) = Path loss in decibels
f = Frequency of transmission
ht = Transmission antenna height
hr = Receiver antenna height
d = Distance between transmitter and receiver

Furthermore, the optimized Okumura-Hata model is defined in Equation 3, as follows:

$$\widehat{PL_{Hata}} = 80.69 + 21.95\,log_{10}(d) + \left[26.16\,log_{10}(f_c) - 3.2(log_{10}(\,11.75h_{ed})\right)^2 - log_{10}(\,h_{gw})(13.82 + 6.55\,log_{10}(\,d))] \quad (3)$$

where,
PL_{Hata} = Estimated path loss using the modified hata model
d = Distance between transmitter and receiver
f_c = Carrier frequency
h_{ed} = Effective receiver antenna height
h_{gw} = Gateway transmitter antenna height

The Close-In (CI) model is grounded in the physical principles of Free Space Path Loss (FSPL), utilizing a reference distance (d_0) of one meter, typically calibrated within an anechoic chamber. The parameter n in the CI model denotes the Path Loss Exponent (PLE), which characterizes the rate of signal attenuation. To derive the optimal n value relative to the empirical field data, an optimization was performed using the Minimum Mean Square Error (MMSE) method. The resulting PLE values obtained for the urban and rural environments are 4.886 and 3.525, respectively. The model is mathematically expressed as [25]:

$$PL^{CI}(f,d) = FSPL(f,d_0) + 10n\,log_{10}\left(\frac{d}{d_0}\right) + X_\sigma^{CI} \quad (4)$$

where,
$PL^{CI}(f,d)$= path loss at frequency f and distance d
FSPL = Free space path loss at reference distance d_0 and frequency f
n = Path loss exponent, a value that depends on the environment
d_0 = Reference distance
X_s = Shadow fading term a random variable representing signal variation due to obstacles

In contrast to the CI model, the Floating-Intercept (FI) model is not anchored to the FSPL reference at a specific distance. Instead, it functions as a purely empirical model that fits a linear regression curve to the logarithmic distance data. The FI model is defined in Equation 5, where the parameter α represents the floating intercept (in dB) and β denotes the slope coefficient, which is analogous to the path loss exponent. Both parameters are derived using the Ordinary Least Squares (OLS) regression method. Similar to the optimization performed for the CI model, the FI parameters were tuned based on field measurements. The optimization for the urban environment yielded values of α = 106.614 and β = 1.498. Meanwhile, the rural environment resulted in values of α = 17.116 and β = 4.095.

$$PL^{FI}(f,d) = \alpha + 10\beta\,log_{10}\left(\frac{d}{d_0}\right) + X_\sigma^{FI} \quad (5)$$

where,
$PL^{FI}(f,d)$= Path loss at frequency f and distance d
α = Floating intercept, a constant term determined empirically
β = Path loss slope factor, represents how fast the path loss increases with distance

The CI model is physically anchored to the free-space path loss at a reference distance and uses the path loss exponent, n, as shown in Table 2., to describe the rate of signal attenuation with distance. A larger n indicates stronger attenuation due to environmental obstruction or scattering. In contrast, the FI model is a purely empirical regression-based model consisting of an intercept parameter, α, and a slope parameter, β. The parameter α represents the floating path loss offset, while β indicates the distance-dependent attenuation trend. Although FI can provide low fitting error, its parameters are less physically constrained than those of the CI model.

Table 2. Fitted parameters of CI and FI models

Model	Environment	Parameter	Value	Physical Meaning
CI	Urban	n	4.886	Path loss exponent in urban morphology
CI	Rural	n	3.525	Path loss exponent in rural morphology
FI	Urban	α	106.614	Empirical floating intercept
FI	Urban	β	1.498	Empirical attenuation slope
FI	Rural	α	17.116	Empirical floating intercept
FI	Rural	β	4.095	Empirical attenuation slope

2.5. Residual Framework

This section outlines the proposed framework for analyzing the residuals specifically, the discrepancies between empirical field measurements and theoretical model estimations. In this study, residuals are not treated merely as random measurement noise. In heterogeneous propagation environments, residuals may contain deterministic components caused by physical and morphological mechanisms that are not explicitly represented in classical models. These mechanisms include multipath reflection from buildings, diffraction around obstacles, vegetation-induced attenuation, and large-scale shadowing.

While residuals are calculated for all baseline propagation models, the optimized Close-In (CI) residual is selected as the primary target for the Machine Learning residual prediction task. This selection was made because the CI model retains physical interpretability through the free-space reference and path loss exponent, while still leaving systematic residual errors under heterogeneous environmental conditions. Mathematically, the residual target is computed using the following equation:

$$e_i = PL_{meas,i} - PL_{CI,opt,i} \quad (6)$$

where e_i is the residual error, $PL_{meas,i}$ denotes the measured path loss value at the i-th point, and $PL_{CI,OPT,i}$ represents the path loss predicted by the optimized CI model. This framework is premised on the hypothesis that if a Machine Learning algorithm can map physically meaningful predictor variables to these residuals with high accuracy, then the model mismatch contains structured and

learnable patterns. The predictor variables utilized encompass the logarithmic distance parameter (X_{dist}) and environmental morphological characteristics (X_{env}), such as area type (urban vs. rural) and obstacle density.

The primary objective of this approach is to evaluate the generalization capability of various Machine Learning algorithms in compensating for the limitations of classical linear models [33]. To assess the effectiveness of these models in capturing residual patterns, two standard statistical metrics were employed: the Root Mean Square Error (RMSE) and the Coefficient of Determination R^2 [34]. The RMSE serves to quantify the magnitude of deviation between the actual residuals and the values predicted by the Machine Learning algorithms. RMSE is calculated using Equation 7 as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_i - \hat{R}_i)^2} \quad (7)$$

Meanwhile, the Coefficient of Determination (R^2) indicates the proportion of residual variance that can be explained by the provided environmental features. A high R^2 value serves as robust evidence that the residuals possess a structured pattern, whereas a value approaching zero implies that the residuals are merely pure random noise that cannot be modeled. The R^2 metric is calculated as follows [34]:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(R_i - \hat{R}_i)^2}{\sum_{i=1}^{N}(R_i - \bar{R})^2} \quad (8)$$

It is crucial to emphasize that the application of Machine Learning algorithms specifically Random Forest, Decision Tree, and Linear Regression in this study is not intended to propose a novel channel model. Rather, Machine Learning is employed as a diagnostic instrument to capture non-linear patterns and deterministic structures embedded within the residuals.

3. Results and Discussion

3.1. Measurement conditions

The anechoic chamber measurement at 1 m produced a path loss of approximately 30 dB. The theoretical FSPL at 915 MHz and 1 m is approximately 31.67 dB. The difference of about 1.67 dB indicates that the measurement setup was reasonably consistent with free-space propagation assumptions, considering practical uncertainties such as antenna alignment, cable loss, hardware tolerance, and RSSI reading fluctuation.

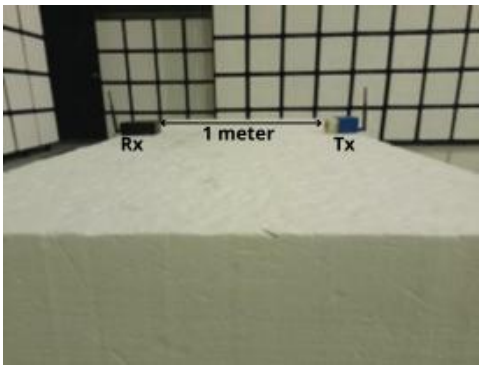

Figure 4. Testing in an anechoic chamber

Figure 5. Testing in anTb urban environment

In the urban area, data collection was conducted along a road under Line-of-Sight (LoS) conditions, as illustrated in Figure 5: Urban Area Testing. As shown in the figure, the testing environment is characterized by high building density and various human activities, ranging from pedestrians and social gatherings to vehicular traffic. The transmitter was positioned on a stationary platform, while the receiver was moved at 10-meter intervals for sequential data acquisition.

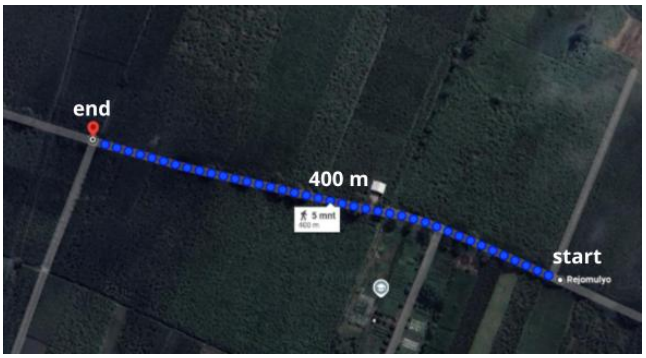

Figure 6. Testing in a rural environment

The second test was conducted in a rural area, where the roadside environment was characterized by dense vegetation and numerous trees. Occasionally, various vehicles ranging from bicycles and motorcycles to trucks passed through the testing site. The positioning of the transmitter and receiver was maintained similarly to the urban area setup. Furthermore, the data acquisition was performed during the same period, specifically during the daytime, to ensure temporal consistency across both environments.

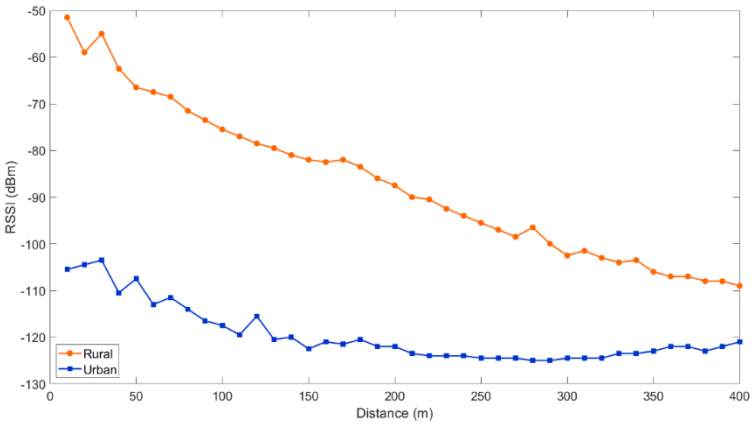

Figure 7. RSSI comparison

As depicted in Figure 7, which illustrates the relationship between RSSI and distance, a significant disparity is evident between the urban and rural environments. In the urban area, the measured RSSI values ranged from a maximum of -105.5 dBm to a minimum of -125 dBm. The comprehensive measurement dataset exhibits a relatively stable rate of decay, characterized by a standard deviation of 6.123. Physical obstructions and multipath reflections within this environment contribute to more substantial signal attenuation. Conversely, in the rural area, the overall measurements yielded a maximum value of -51.5 dBm and a minimum of -109 dBm. The data indicates a gradual decline with a standard deviation of 16.150, which is notably higher than that observed in the urban setting. This finding suggests that in rural conditions, signal propagation is superior due to the scarcity of obstacles such as walls, partitions, or other dense materials.

The RSSI results indicate that environmental morphology plays a stronger role than distance alone in LoRa propagation. The urban route produced consistently lower RSSI values because the propagation path was surrounded by buildings, vegetation, vehicles, and human activities. Although the measurements were conducted under LoS conditions, reflected and diffracted signal components may still interfere with the dominant signal path. This interpretation is consistent with recent LoRa propagation studies showing that local morphology, building density, and vegetation can significantly affect RSSI stability and signal attenuation. Azevedo and Mendonça emphasized that LoRa propagation models must be selected according to deployment environment because model accuracy is highly site-dependent [19]. Similar behavior was also reported by Anisah et al., who found that LoRa propagation in campus environments requires local model adjustment due to the influence of surrounding structures [17].

In contrast, the rural route showed higher RSSI values because the propagation path was more open and had fewer solid obstructions. However, the rural RSSI had a larger standard deviation than the urban RSSI. This shows that better average signal strength does not always mean more stable propagation. Vegetation, terrain variation, and occasional vehicle movement can still cause local RSSI fluctuation. This finding is supported by Mahjoub et al., who reported that LoRa signal propagation in agricultural areas is strongly affected by vegetation density and site-specific morphology [21]. Therefore, the rural condition in this study should be interpreted as an open but still environmentally dynamic propagation route.

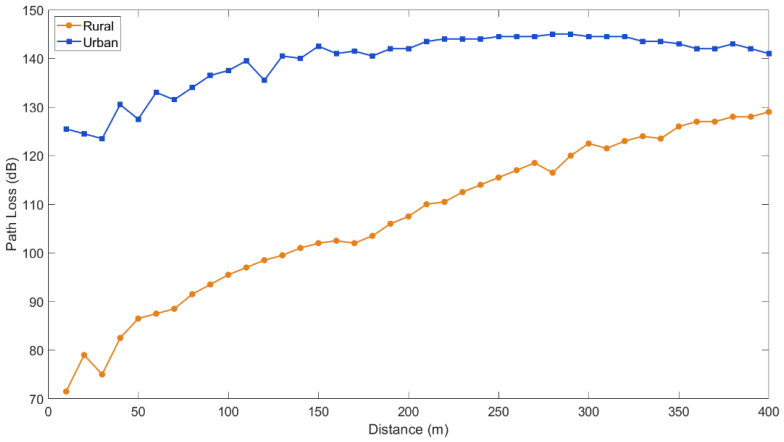

Figure 8. Path loss comparison

The measurement results presented in Figure 8 demonstrate a distinct divergence in path loss values between the urban and rural environments. At a distance of 10 meters, the path loss in the urban area reached approximately 125.5 dB, whereas in the rural area it was only about 71.5 dB, indicating a signal attenuation differential of ±54 dB.As the distance increases, path loss in both

environments rises, yet the growth patterns differ significantly. In the urban area, the increase is relatively rapid up to a distance of ±150 meters, reaching a value of approximately 144.5 dB, after which it tends to stabilize up to 400 meters, fluctuating within the 141–145 dB range. Conversely, the rural area exhibits a more consistent and linear increase, starting from 71.5 dB at 10 meters and rising to 129 dB at 400 meters, with no indication of saturation.

The path loss pattern confirms that the urban environment introduced stronger attenuation than the rural environment. At short distances, the difference between urban and rural path loss was already large, indicating that the surrounding morphology affected the signal even before distance became the dominant factor. In the urban route, path loss increased rapidly at early distances and then tended to saturate after approximately 150 m. This saturation suggests that the received signal was already close to the sensitivity limit of the receiver, making additional distance less visible in the measured RSSI variation. Such behavior is commonly observed in obstructed LoRa deployment, where local scattering, diffraction, and shadowing can dominate over the ideal log-distance trend.

The rural path loss increased more gradually and followed a more distance-dependent trend. This result agrees with Lima et al., who showed that LoRa propagation in open and vegetation-dominated regions is strongly affected by local morphology, but the path loss trend can remain more progressive than in dense urban conditions [20]. A similar observation was reported by Xu et al. in agricultural LoRa measurements, where crop density and vegetation structure influenced signal attenuation along the measurement route [34]. Therefore, the difference between the urban and rural curves in this study indicates that LoRa path loss cannot be generalized using distance alone, even when both measurements are conducted under LoS conditions.

Several sources of uncertainty may affect the field measurements, including RSSI fluctuation, antenna orientation, hardware sensitivity tolerance, temporary obstruction from surrounding activities, and environmental scattering. In the urban environment, moving objects, buildings, and vegetation may introduce additional multipath and shadowing variations. In the rural environment, vegetation and roadside vehicle movement may also contribute to local signal fluctuations. These uncertainties explain why residual errors remain even after propagation model optimization.

Table 3. Simple statistical results

Environment	Mean RSSI	Std RSSI	Mean Path Loss	Std Path Loss
Rural	-87.112 dBm	16.151	107.112	16.151
Urban	-119.538 dBm	6.123	139.538	6.123

In addition to the descriptive statistics shown in Table 3, Welch's t-test was conducted to examine whether the path loss difference between urban and rural environments was statistically significant. The result showed a significant difference, with $p < 0.001$. This confirms that the urban and rural measurement conditions produced statistically different propagation characteristics. The statistical result strengthens the visual interpretation of the RSSI and path loss curves. The significant difference between urban and rural path loss indicates that the two environments represent different propagation regimes rather than random measurement variations.

The mean urban path loss of 139.538 dB was substantially higher than the rural mean path loss of 107.112 dB. This difference supports the argument that urban obstruction produces stronger large-scale attenuation. However, the larger rural standard deviation indicates that rural propagation was not completely uniform. This finding is important because it shows that rural LoRa planning should also consider vegetation and terrain variability, not only open-space distance.

Compared with previous studies, the results support the general conclusion that LoRa propagation is highly environment-specific. Previous LoRa studies in urban and campus environments reported that buildings and local infrastructure can significantly change RSSI and path loss behavior [17], [19]. Meanwhile, studies in agricultural and vegetation environments reported that vegetation density can produce additional attenuation and fluctuation [20], [21], [34]. The present study extends these findings by showing both conditions in a single measurement framework and confirming the difference statistically using the same hardware configuration, frequency, and measurement interval.

3.2. Comparison of Classical Propagation Models

The optimized FI model in Table 4., achieved the lowest RMSE in both urban and rural environments, indicating that a flexible empirical intercept and slope can effectively fit the measured path loss trend. However, the low error of FI should be interpreted carefully because its parameters are purely empirical and less physically constrained than those of the CI model. The optimized CI model preserved stronger physical meaning through the path loss exponent, but it still produced considerable residuals, particularly in the rural environment. This indicates that a single path loss exponent may be insufficient to represent environmental variations along the measurement route.

Table 4. RMSE residual model

Model	Urban RMSE (dB)	Rural RMSE (dB)
Optimized CI	13.066	4.877
Optimized FI	2.253	4.370
Standard CI	49.003	17.697
Standard FI	76.948	45.750
Okumura-Hata	108.076	139.971
Optimized Okumura-Hata	63.295	95.990

The RMSE comparison shows that the optimized FI model achieved the lowest error in both environments, with 2.253 dB in the urban route and 4.370 dB in the rural route. This result indicates that the FI model can closely follow the measured path loss curve because its intercept and slope are fully estimated from field data. This behavior is consistent with previous reviews showing that empirical models can provide good fitting accuracy when calibrated using local measurements [19]. However, this advantage should be interpreted carefully because the FI parameters are less physically constrained than those of the CI model. Therefore, FI is suitable for local curve fitting, but it may have limited interpretability when applied to different environments.

The optimized CI model produced higher RMSE than the optimized FI model, particularly in the urban environment. This difference occurs because the CI model is anchored to FSPL and relies on the path loss exponent to represent distance-dependent attenuation. The model is physically meaningful, but a single path loss exponent cannot fully represent sudden morphology changes caused by buildings, vegetation, and roadside objects. Similar conclusions were reported in recent LoRa propagation studies, where local calibration was required to improve model accuracy in specific deployment environments [19], [20]. Thus, the CI model remains useful as a physically interpretable baseline, but residual analysis is needed to understand the remaining mismatch.

The Okumura-Hata model produced the largest RMSE in both environments. This result is reasonable because Okumura-Hata was originally developed for macrocellular scenarios and does not

directly represent low-height, short-to-medium-range LoRa links. The failure of this model in the present study does not mean that Okumura-Hata is invalid in all cases, but it shows that its original assumptions are not compatible with the tested LoRa scenario. This finding is also consistent with recent reviews of LoRa propagation models, which emphasize that empirical models must be carefully selected and validated for each operating environment [19]. Therefore, the high RMSE of Okumura-Hata confirms the need for local model evaluation before applying classical propagation models to LoRa radio planning.

3.3. Residual Analysis

Autocorrelation analysis in Table 5., was performed to examine whether the residuals were randomly distributed or contained distance-dependent structure. The optimized CI residuals exhibited strong autocorrelation, particularly at lag-1, with values of approximately 0.974 in the urban environment and 0.894 in the rural environment. This indicates that residual values at adjacent measurement distances were strongly related. Therefore, the residuals cannot be interpreted as purely random noise. Instead, they exhibit structured behavior associated with gradual environmental changes along the propagation path. Since the autocorrelation analysis already demonstrated structured residual dependency, FFT analysis was considered optional and was not emphasized in this study.

Table 5. Autocorrelation analysis

Environment	Residual Target	ACF lag-1	ACF lag-2	ACF lag-3	ACF lag-4	ACF lag-5
Urban	gapOptimizedCI	0.974	0.959	0.952	0.926	0.929
Rural	gapOptimizedCI	0.894	0.701	0.702	0.559	0.333

The autocorrelation result provides stronger evidence that the optimized CI residuals are not purely random. The lag-1 autocorrelation values of 0.974 in the urban environment and 0.894 in the rural environment indicate that adjacent residual points were strongly related. This means that the residual error changed gradually along the measurement route instead of fluctuating independently at each point. In radio propagation, this behavior can be associated with large-scale environmental effects, such as shadowing, gradual changes in obstacle density, and terrain-related attenuation. Therefore, the residual pattern represents an environmental structure that is not fully captured by the CI path loss exponent.

This result is relevant to recent machine learning-based radio propagation studies. Vasudevan and Yuksel explained that machine learning can capture nonlinear relationships between propagation loss and environmental features that are difficult to represent using classical analytical models [22]. In this study, autocorrelation analysis supports the same interpretation from a residual perspective. The residuals are not treated as unexplained noise, but as remaining information that can still be modeled using distance, environment type, and obstacle density. This strengthens the novelty of the study because the analysis moves beyond RMSE comparison and examines the internal structure of model error.

3.4. Performance of Machine Learning in Residual Prediction

Following the characterization of residuals in both urban and rural environments, the subsequent phase involves evaluating the efficacy of Machine Learning algorithms in discerning the underlying residual structures. This evaluation entails a comparative analysis of three distinct algorithms: Random Forest (RF), Decision Tree (DT), and Linear Regression (LR). The assessment is based on the Coefficient of Determination (R^2) and Root Mean Square Error (RMSE) metrics, as summarized in Table 6.

Table 6. Comparison of machine learning algorithms for residual prediction

No	Algorithm	Mean CV R^2	Mean CV RMSE
1	Linear Regression	0.865	6.307
2	Decision Tree	0.959	3.515
3	Random Forest	0.972	2.704

The machine learning results show that nonlinear models were more effective than Linear Regression in predicting optimized CI residuals. Linear Regression achieved a mean CV R^2 of 0.865, indicating that part of the residual variation still followed a linear tendency. However, Decision Tree and Random Forest achieved higher R^2 values of 0.959 and 0.972, respectively. This improvement indicates that residual behavior was influenced by nonlinear interactions among logarithmic distance, environment type, and obstacle density. The result supports the assumption that residuals contain learnable environmental patterns rather than only random measurement error.

The superior performance of Random Forest is consistent with recent studies showing that machine learning can improve radio propagation modeling by learning nonlinear and site-specific effects. Vasudevan and Yuksel highlighted that ensemble and tree-based models are useful when propagation loss depends on complex environmental features [22]. Alobaidy et al. also showed that hybrid machine learning can improve LoRa-based IoT propagation characterization by incorporating environmental and measurement-related variables [7]. However, the high performance in this study should be interpreted with caution because the dataset consists of 80 samples. Therefore, cross-validation was used to reduce evaluation bias, and future work should validate the residual model using larger datasets and additional LoS/NLoS scenarios.

The learning curve in Figure 9 shows that the validation R^2 score increased as the number of training samples increased, while the training R^2 score remained consistently high. This trend indicates that the Random Forest model benefited from additional data and that the residual prediction task was not merely based on memorization. Nevertheless, the remaining gap between training and validation performance suggests that the model is still sensitive to dataset size and requires more data to improve generalization. Therefore, Random Forest in this study should not be interpreted as a universal propagation model, but rather as a diagnostic tool showing that the residuals of classical propagation models still contain structured information related to environmental conditions. Future studies with larger and more diverse datasets are required to reduce this gap and improve the robustness of the residual-based prediction framework.

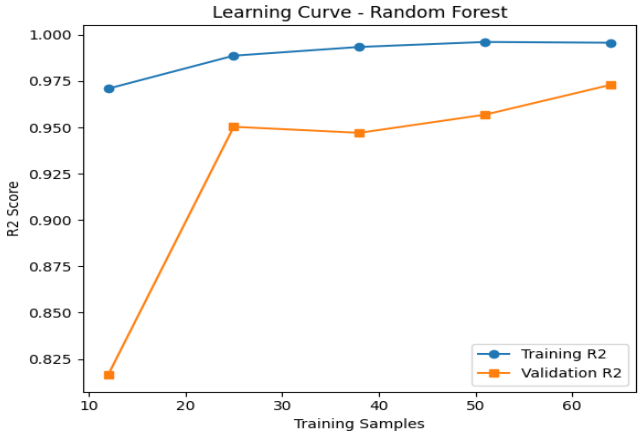

Figure 9. Learning curve of Random Forest residual prediction

To clarify the position of the present study among previous works, Table 7 compares several recent LoRa propagation and machine learning-based radio propagation studies with the findings of this research. The comparison highlights the measurement condition, main findings, and the specific contribution of the present study.

Table 7. Comparison between previous studies and the present study

Reference	Study Condition / Environment	Main Findings	Relation to the Present Study
Anisah et al. [17]	LoRa propagation measurement in a campus area	The study showed that campus infrastructure affects LoRa propagation and requires local model adjustment.	Supports the urban campus result in this study, where buildings and surrounding structures caused lower RSSI and higher path loss.
Azevedo and Mendonça [19]	Review of LoRa propagation models in different environments	The study emphasized that LoRa model accuracy is strongly dependent on deployment environment and model selection.	Supports the need to evaluate CI, FI, and Okumura-Hata models separately for urban and rural environments.
Lima et al. [20]	LoRaWAN 915 MHz propagation modeling in Amazon regions	Local morphology was shown to strongly affect RSSI and path loss behavior.	Supports the finding that rural and vegetation-related morphology influences path loss variation.
Mahjoub et al. [21]	LoRa propagation in agricultural/oasis environments	Vegetation density and site characteristics affected LoRa signal attenuation.	Supports the interpretation that rural RSSI fluctuation can occur due to vegetation and terrain variation.
Vasudevan and Yuksel [22]	Machine learning for radio propagation modeling	Machine learning can capture nonlinear relationships between propagation loss and environmental features.	Supports the use of ML to analyze nonlinear residual patterns rather than treating residuals as random noise.
Alobaidy et al. [7]	LoRa-based IoT propagation characterization using hybrid machine learning	Machine learning improved LoRa propagation characterization by using measurement and environmental variables.	Supports the Random Forest result, which achieved the best residual prediction in this study.
Xu et al. [34]	LoRa propagation in agricultural crop environments	Vegetation structure influenced LoRa attenuation along the measurement route.	Supports the discussion that rural LoRa propagation is not purely free-space because vegetation still affects signal behavior.
Present study	Urban campus environment at Universitas Negeri Surabaya and rural area in Kediri, Indonesia, using 915 MHz LoRa with 80 LoS samples	Urban measurements produced higher path loss than rural measurements. Optimized FI achieved the lowest RMSE, while optimized CI residuals showed strong autocorrelation. Random Forest achieved $R^2 = 0.972$ and RMSE = 2.704 dB.	This study extends previous work by not only comparing propagation models, but also analyzing residual errors as structured environmental information using autocorrelation and machine learning.

As shown in Table 7, previous studies have demonstrated that LoRa propagation is strongly affected by deployment environment, vegetation, and local morphology. However, most of them focused on direct path loss modeling or model accuracy comparison. The present study extends these works by analyzing the residual structure of classical propagation models and showing that the remaining model error contains learnable environmental patterns.

Overall, the results indicate that LoRa propagation mismatch is not caused only by measurement noise or random interference. The urban and rural measurements showed different RSSI, path loss,

RMSE, residual autocorrelation, and machine learning behavior. The comparison with previous studies confirms that LoRa propagation is highly site-dependent, while the present findings add that the remaining error of a physically grounded model can still be analyzed and predicted. Therefore, the main contribution of this study is not only the comparison of classical models, but also the interpretation of residuals as structured environmental information. This provides a practical direction for future LoRa radio planning, where classical models can be combined with residual-based compensation to improve prediction reliability in heterogeneous environments.

4. Conclusion

This study demonstrates that LoRa propagation behavior differs significantly between urban and rural environments. The urban route produced higher path loss, with a mean value of 139.538 dB, compared with 107.112 dB in the rural route. The optimized FI model achieved the lowest baseline RMSE, but its empirical nature limits physical interpretation. In contrast, the optimized CI model retained stronger physical meaning but still produced structured residuals, indicating that distance-based parameters alone are insufficient to represent heterogeneous propagation effects.

The residual analysis confirms that the mismatch between measured path loss and classical model prediction is not purely random. The optimized CI residuals showed strong lag-1 autocorrelation, with values of 0.974 in urban and 0.894 in rural environments. Random Forest provided the best residual prediction, achieving $R^2 = 0.972$ and RMSE = 2.704 dB. These results indicate that logarithmic distance, environment type, and obstacle density can explain a substantial part of the remaining model error. Therefore, residual-based machine learning can serve as a useful diagnostic and compensation approach for LoRa radio planning in heterogeneous IoT environments.

References

[1] Sadeghi-Niaraki, A. (2023). Internet of Thing (IoT) review of review: Bibliometric overview since its foundation. *Future Generation Computer Systems*, *143*, 361-377.

[2] J. Linder, (2026). Iot Industry Statistics, Gitnux. [Online]. Available: https://gitnux.org/iot-industry-statistics/

[3] Duguma, A. L., & Bai, X. (2024). How the internet of things technology improves agricultural efficiency. *Artificial Intelligence Review*, *58*(2), 63.

[4] Sutanto, E., Fahmi, F., Nurdin, Y., Puspitaningayu, P., & Aziz, M. (2024). Microgrid monitoring for patient safety in hospital using distributed residual current. *IEEE Internet of Things Journal*, *11*(18), 29979-29992.

[5] Hassija, V., Chamola, V., Saxena, V., Jain, D., Goyal, P., & Sikdar, B. (2019). A survey on IoT security: application areas, security threats, and solution architectures. *IEEE access*, *7*, 82721-82743.

[6] Manzano, L. G., Boukabache, H., Danzeca, S., Heracleous, N., Murtas, F., Perrin, D., ... & Silari, M. (2021). An IoT LoRaWAN network for environmental radiation monitoring. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-12.

[7] Alobaidy, H. A., Abdullah, N. F., Nordin, R., Behjati, M., Abu-Samah, A., Maizan, H., & Mandeep, J. S. (2024). Empowering extreme communication: propagation characterization of a LoRa-based internet of things network using hybrid machine learning. *IEEE Open Journal of the Communications Society*, *5*, 3997-4023.

[8] Lloret, J., Garcia, L., Jimenez, J. M., Sendra, S., & Lorenz, P. (2021). Cluster-based communication protocol and architecture for a wastewater purification system intended for irrigation. *IEEE Access*, *9*, 142374-142389.

[9] Suji Prasad, S. J., Thangatamilan, M., Suresh, M., Panchal, H., Rajan, C. A., Sagana, C., ... & Sadasivuni, K. K. (2022). An efficient LoRa-based smart agriculture management and

monitoring system using wireless sensor networks. *International Journal of Ambient Energy*, *43*(1), 5447-5450.

[10] Dallalbashi, Z. E., Alhayalir, S., Mnati, M. J., & Alhayali, A. A. A. (2023). Low-cost battery monitoring circuit for a photovoltaic system based on LoRa/LoRaWAN network. *Indonesian Journal of Electrical Engineering and Computer Science*, *29*(2), 669-677.

[11] Ahmad, K. A., Segaran, J. D., Hashim, F. R., & Jusoh, M. T. (2017). Lora propagation at 433 MHz in tropical climate environment. *Journal of Fundamental and Applied Sciences*, *9*(3S), 384-394.

[12] Urabe, I., Li, A., Fujisawa, M., Kim, S. J., & Hasegawa, M. (2023). Combinatorial mab-based joint channel and spreading factor selection for lora devices. *Sensors*, *23*(15), 6687.

[13] Hudiono, H., Taufik, M., Perdana, R. H. Y., & Rakhmania, A. E. (2021). Digital centralized water meter using 433 MHz LoRa. *Bulletin of Electrical Engineering and Informatics*, *10*(4), 2062-2071.

[14] Anzum, R. (2021). Factors that affect LoRa propagation in foliage medium. *Procedia Computer Science*, *194*, 149-155.

[15] Paredes, M., Bertoldo, S., Carosso, L., Lucianaz, C., Marchetta, E., Allegretti, M., & Savi, P. (2019). Propagation measurements for a LoRa network in an urban environment. *Journal of Electromagnetic Waves and Applications*, *33*(15), 2022-2036.

[16] Săcăleanu, D. I., Manciu, I. P., & Perişoară, L. A. (2019, June). Performance analysis of LoRa technology in wireless sensor networks. In *2019 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)* (pp. 1-5). IEEE.

[17] Anisah, I., & Yuliana, M. (2023, December). Experimental results of LoRa network radio propagation modeling in campus area. In *2023 6th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)* (pp. 427-432). IEEE.

[18] El Chall, R., Lahoud, S., & El Helou, M. (2019). LoRaWAN network: Radio propagation models and performance evaluation in various environments in Lebanon. *IEEE Internet of Things Journal*, *6*(2), 2366-2378.

[19] Azevedo, J. A., & Mendonça, F. (2024). A critical review of the propagation models employed in LoRa systems. *Sensors*, *24*(12), 3877.

[20] Lima, W. G., Lopes, A. V., Cardoso, C. M., Araújo, J. P., Neto, M. C., Tostes, M. E., ... & Barros, F. J. (2024). LoRa technology propagation models for IoT network planning in the Amazon regions. *Sensors*, *24*(5), 1621.

[21] Mahjoub, T., Mnaouer, A. B., Said, M. B., & Boujemaa, H. (2024). LoRa signal propagation and path loss prediction in Tunisian date palm oases. *Computers and Electronics in Agriculture*, *222*, 109027.

[22] Vasudevan, M., & Yuksel, M. (2024). Machine learning for radio propagation modeling: A comprehensive survey. *IEEE Open Journal of the Communications Society*, *5*, 5123-5153.

[23] Masadan, N. A. B., Habaebi, M. H., & Yusoff, S. H. (2018, September). LoRa LPWAN propagation channel modelling in IIUM campus. In *2018 7th International Conference on Computer and communication Engineering (ICCCE)* (pp. 14-19). IEEE.

[24] Alan, M. B. F., Puspitaningayu, P., Sutanto, E., Fahmi, F., & Widayaka, P. D. (2024, January). Design and implementation of photovoltaic monitoring system using Arduino UNO R4. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 899-903). IEEE.

[25] Puspitaningayu, P., Alan, M. B. F., Sutanto, E., Fahmi, F., & Kartini, U. T. (2024, January). An Evaluation of LoRaWAN and WLAN for IoT-based Photovoltaic Microgrid Monitoring. In *2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)* (pp. 1255-1259). IEEE.

[26] Díaz Zayas, A., Rivas Tocado, F. J., & Rodríguez, P. (2020). Evolution and testing of NB-IoT solutions. *Applied Sciences*, *10*(21), 7903.
[27] Marini, R., Mikhaylov, K., Pasolini, G., & Buratti, C. (2022). Low-power wide-area networks: Comparison of LoRaWAN and NB-IoT performance. *IEEE Internet of Things Journal*, *9*(21), 21051-21063.
[28] Al-Shareeda, M. A., Alsadhan, A. A., Qasim, H. H., & Manickam, S. (2023). Long range technology for internet of things: review, challenges, and future directions. *Bulletin of Electrical Engineering and Informatics*, *12*(6), 3758-3767.
[29] Budi, B., & Machdi, A. R. (2024). Distance Testing on Point to Point Communication with Lora Based on RSSI and Log Normal Shadowing Model. *Journal Of Energy And Electrical Engineering*, *5*(2).
[30] de Camargo, E. T., Spanhol, F. A., & Castro e Souza, Á. R. (2021). Deployment of a LoRaWAN network and evaluation of tracking devices in the context of smart cities. *Journal of Internet Services and Applications*, *12*(1), 8.
[31] Zhang, H., Meng, F., Xu, J., Liu, Z., & Meng, J. (2022). Evaluation of machine learning models for daily reference evapotranspiration modeling using limited meteorological data in Eastern Inner Mongolia, North China. *Water*, *14*(18), 2890.
[32] Mahjoub, T., Mnaouer, A. B., Said, M. B., & Boujemaa, H. (2024). LoRa signal propagation and path loss prediction in Tunisian date palm oases. *Computers and Electronics in Agriculture*, *222*, 109027.
[33] Iliev, I., Velchev, Y., Petkov, P. Z., Bonev, B., Iliev, G., & Nachev, I. (2024). A machine learning approach for path loss prediction using combination of regression and classification models. *Sensors*, *24*(17), 5855.
[34] Xu, T., Ma, D., Fang, W., & Huang, Y. (2025). Experimental study on the propagation characteristics of LoRa signals in maize fields. *Electronics*, *14*(11), 2156.