JURNAL INFOTEL
VOL. 18, NO. 4, NOVEMBER 2026, PP. 577–600. DOI:10.20895/INFOTEL.V18I4.10944

RESEARCH ARTICLE

A Hybrid Ensemble Learning Framework for Path Loss Prediction in River Propagation Environments

Bengawan Alfaresi[1,*], Feby Ardianto[2], Sofiah[3], and Bungkus Aji Daya[4]

[1,2,3,4]Electrical Engineering Study Program, Faculty of Engineering, Universitas Muhammadiyah Palembang 30263, Indonesia

*Corresponding email: begawan_alfarezi@um-palembang.ac.id

Received: April 26, 2026; Revised: October 13, 2026; Accepted: October 30, 2026.

Abstract: Accurate path-loss prediction is essential for reliable wireless-network planning, particularly in riverine environments where over-water propagation, shoreline geometry, reflections, diffraction, and heterogeneous surrounding structures can reduce the accuracy of conventional empirical models. This study investigates a hybrid ensemble-learning framework for path-loss prediction using 9,761 measurement samples collected along a 13.75 km route on the Musi River, Indonesia, at 1800 and 2100 MHz. We used Random Forest (RF) and eXtreme Gradient Boosting (XGBoost) as base learners. We combined them using two ensemble strategies: weighted blending and stacking, with Support Vector Regression (SVR) as the meta-learner in the final stacking architecture. Model hyperparameters were optimized using GridSearchCV, while predictive performance was evaluated using 10-fold cross-validation and assessed through mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination R^2. The proposed models were compared with individual machine-learning models and conventional empirical propagation models. Among the empirical baselines, COST-231 performed best, with an MAE of 10.7504 dB, RMSE of 13.1797 dB, and R^2 of 0.1721. The stacking hybrid produced the best overall results, achieving an MAE of 2.6213 dB, RMSE of 3.6885 dB, and R^2 of 0.9352, corresponding to reductions of approximately 75.6% in MAE and 72.0% in RMSE relative to COST-231. These findings demonstrate that the stacking-based ensemble provides a more accurate and practical data-driven solution for path-loss prediction in complex riverine propagation environments.

Keywords: Ensemble learning; Path loss prediction; Riverine propagation; Stacking regression; Weighted blending

© Jurnal Infotel ISSN: 2085-3688; e-ISSN: 2460-0997

1 Introduction

Wireless communication has evolved into a fundamental technological infrastructure that enables numerous contemporary services, from public-sector operations and industrial activities to increasingly intensive access to digital information [1, 2]. As wireless networks continue to expand, their deployment is no longer confined to conventional urban environments but increasingly extends to rural, indoor, transportation, and water-associated areas [3]. Communication between transmitting and receiving nodes relies on electromagnetic-wave propagation through the surrounding medium rather than on physical transmission cables [4,5]. Consequently, propagation performance is strongly affected by the characteristics of the environment through which the signal travels. Terrain configuration, vegetation, buildings, and water surfaces can modify signal attenuation and ultimately influence the power received at the terminal [6]. Path loss is therefore commonly used to quantify the reduction in signal strength between a transmitter and a receiver [7, 8]. Accurately estimating this attenuation is fundamental to planning, optimizing, and deploying wireless communication systems [7, 9, 10].

In radio-network engineering, path-loss estimates provide essential information for determining coverage areas, base-station locations, operating frequencies, and link-budget requirements [10, 11]. Their importance has grown with the rise of Internet of Things (IoT) applications, where large numbers of low-power devices must maintain reliable connectivity under strict energy constraints [12–15]. Accurate knowledge of propagation loss can help network designers select transmission power levels, packet-transmission intervals, retransmission mechanisms, and operating channels [16, 17]. Such decisions are particularly important in environmental monitoring and distributed sensing applications located in coastal, maritime, agricultural, and other geographically challenging areas. Appropriate propagation estimates can consequently improve communication reliability, extend battery lifetime, and reduce network operating requirements [13, 18].

Conventional network planning has traditionally relied on empirical propagation formulations, including Hata, Okumura, COST-231, and Egli, because these models are computationally simple and straightforward to apply [4, 14]. In particular, the Hata and Okumura formulations have been extensively employed for terrestrial urban and suburban deployments [6]. However, empirical models are generally constructed from measurement campaigns performed under specific environmental conditions and are expressed through predefined mathematical relationships. Their predictive reliability can therefore deteriorate when the actual propagation environment differs substantially from the conditions represented during model development [8, 11]. This limitation becomes increasingly relevant in areas characterized by transitions between land and water, irregular shoreline structures, and highly heterogeneous propagation geometries.

Propagation over water differs from that typically encountered in terrestrial communication systems [19, 20]. Water surfaces can generate strong reflected components, while changes in the relative geometry of the transmitter, receiver, shoreline, and water surface can alter the phase and amplitude of received multipath signals [19]. Environmental changes such as water-level variation can further modify effective propagation geometry and reflection conditions [21]. In river corridors, the situation is additionally complicated by irregular channel shapes, shoreline vegetation, nearby buildings, and repeated transitions between open-water and land-dominated propagation. These factors can create constructive and destructive interference, shadowing, diffraction, and localized scattering

effects that are difficult to describe using a single fixed-form propagation equation. These characteristics motivate models that can adapt to complex relationships learned directly from measurement data.

Artificial intelligence and machine-learning methods have consequently gained increasing attention as alternatives for radio-propagation prediction [12, 22]. Algorithms such as Random Forest (RF), Support Vector Regression (SVR), and eXtreme Gradient Boosting (XGBoost) have shown strong predictive capability in environments with nonlinear, multivariable propagation behavior [2, 23, 24]. Rather than relying exclusively on predetermined analytical relationships, machine-learning algorithms infer relationships among propagation variables from observed data. This ability lets them incorporate geometrical and environmental descriptors simultaneously and represent interactions that may be difficult to capture with conventional empirical formulations [25].

Although machine learning has been widely explored for urban, terrestrial, and related wireless scenarios [26], comparatively fewer studies have addressed environments dominated by rivers or mixed land–water propagation [25, 27]. One key challenge is acquiring representative measurement data. Field measurements over rivers and other aquatic environments generally involve greater logistical complexity, higher measurement costs, and stronger dependence on environmental conditions than standard terrestrial campaigns [21, 28]. These constraints can limit both the amount and diversity of available training data. Furthermore, an individual machine-learning algorithm's predictive behavior may depend strongly on the dataset's characteristics and the selected model configuration. Ensemble learning offers one approach to address this limitation by combining predictions from multiple learners, exploiting complementary strengths [29, 30]. Stacking and weighted ensemble strategies have also delivered consistent accuracy gains in other prediction domains, including monthly rainfall forecasting [31], ensemble forecasting of geophysical events [32], energy-performance prediction of residential buildings [33], and forecasting of nonlinear dynamics containing extreme events [34].

Recent studies increasingly support ensemble strategies for radio-propagation prediction. Elmezughi et al. [8] developed an ensemble framework combining different neural-network architectures and demonstrated improved path-loss and shadowing estimation compared with conventional modeling approaches. Kwon and Son [7] similarly reported that combining optimized neural networks could improve path-loss prediction performance relative to individual predictors. Afape et al. [30] later investigated automated, hyperparameter-tuned stacking regression for millimeter-wave path-loss estimation and showed that stacking could provide competitive improvements over empirical and standalone machine-learning models. Collectively, these findings indicate that ensemble learning can improve predictive robustness by exploiting information from multiple component models. Nevertheless, most recent ensemble studies have been conducted in terrestrial, indoor, or millimeter-wave environments, leaving water-dominated and riverine propagation comparatively underinvestigated.

Recent developments in maritime wireless research also demonstrate increasing use of data-driven models for water-related propagation problems. A 2025 study on land-to-ship wireless links compared several machine-learning and regression approaches for predicting received signal characteristics in maritime environments [35]. In 2026, Ding et al. introduced a hybrid ConvLSTM–PSO–XGBoost framework for maritime 5G signal-strength prediction, illustrating the potential of combining complementary learning mechanisms in complex maritime conditions [36]. Another measurement-driven maritime study pub-

lished in 2026 evaluated multiple machine-learning approaches, including ensemble-based techniques, for shore-to-ship path-loss modeling [37]. These studies extend data-driven propagation research toward water-associated scenarios; however, maritime channels and river corridors do not necessarily exhibit identical propagation characteristics. River environments generally involve narrower water bodies, stronger interaction with nearby land structures, irregular riverbank morphology, and greater influence from surrounding buildings and vegetation.

Several earlier studies have also examined path-loss prediction using machine-learning and deep-learning techniques. The work in [8] combined ANN, RNN-LSTM, and CNN-based predictors in an ensemble framework and reported greater modeling flexibility than conventional linear path-loss and log-normal shadowing formulations. The study in [4] investigated a Bidirectional Long Short-Term Memory approach and demonstrated improved path-loss estimation relative to traditional benchmark models. For mixed water–land propagation, the study in [38] integrated evolutionary feature-selection techniques, including genetic algorithm, particle swarm optimization, and univariate selection, with machine-learning predictors; the univariate–Random Forest configuration performed best among the evaluated alternatives. In another water-related propagation problem, [39] employed a deep neural network to estimate path loss under evaporation-duct conditions and demonstrated high prediction accuracy. However, model behavior varied across operating frequencies. Artificial-intelligence-based path-loss prediction at 28 GHz for an indoor 5G scenario was also investigated in [40]. Taken together, these studies demonstrate significant progress in data-driven propagation modeling, while also showing that the investigated environments, model structures, and propagation conditions remain highly diverse.

Despite this progress, a specific research gap remains in riverine radio-propagation modeling. Existing studies involving mixed land–water or river-related measurements have predominantly emphasized individual machine-learning algorithms, feature-selection procedures, deep-learning models, or conventional propagation formulations. Meanwhile, recent ensemble-oriented studies [7, 8, 30] and emerging maritime investigations [35–37] demonstrate the potential value of combining multiple predictive models, but systematic evaluation of different ensemble architectures for river-specific path-loss prediction remains limited. In particular, direct comparisons of weighted blending and stacking against both conventional empirical models and standalone machine-learning approaches under the same riverine measurement conditions have received little attention. Addressing this gap matters because river corridors involve over-water propagation, shoreline interactions, land–water transitions, and heterogeneous structures that may produce propagation patterns not adequately represented by conventional terrestrial models.

Previous studies using riverine or mixed land–water measurements have primarily focused on individual machine-learning models, feature-selection strategies, or conventional propagation modeling. However, a systematic comparison of hybrid ensemble architectures, particularly weighted blending and stacking, against both empirical propagation models and individual machine-learning baselines remains limited for riverine path-loss prediction. This gap motivates the present study.

This study investigates hybrid ensemble learning for path-loss prediction in a riverine environment using measurement data collected along the Musi River, Indonesia. Random Forest (RF) and XGBoost are employed as base learners, while weighted blending and stacking are evaluated as alternative ensemble strategies. This work is primarily comparative and application-oriented rather than algorithmic. Specifically, the study systematically

compares conventional empirical propagation models, individual machine-learning models, and hybrid ensemble approaches under the same measurement conditions. In addition, feature-importance analysis is incorporated to examine the relative predictive contribution of conventional propagation parameters and river-related spatial descriptors. Through this evaluation, the study aims to determine whether ensemble learning can provide a more accurate and interpretable solution for path-loss prediction in riverine propagation environments.

2 Research Method

2.1 Data Collection and Measurement Configuration

The field measurement campaign was conducted using a portable measurement setup consisting of a laptop computer, mobile device, GPS receiver, and USB hub. These components were integrated to support signal data acquisition, geographic positioning, and measurement logging during the drive-test campaign. The configuration of the field measurement devices is illustrated in Figure 1.

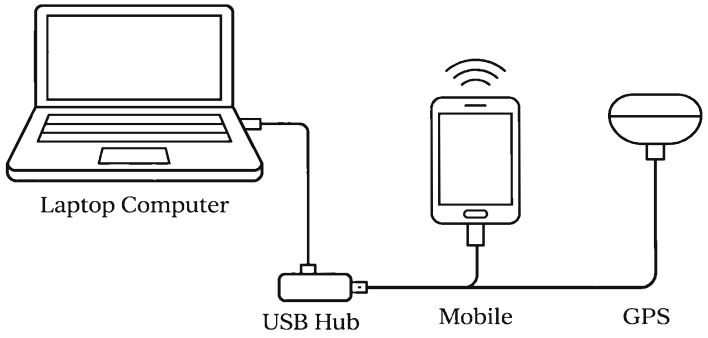


Figure 1: Field measurement devices.

Figure 1 illustrates the configuration of the field measurement devices used during data acquisition. The smartphone served as the primary measurement terminal and was connected to a GPS receiver to obtain geographical coordinates for each measurement point. The laptop computer was used to operate the measurement and monitoring software, while the USB hub and data cables provided the physical connection among the devices. This configuration enabled simultaneous recording of signal measurements and location information during the measurement campaign. The hardware, software, and supporting tools used in the field measurement process are summarized in Table 1.

Environmental characteristics such as river width, current velocity, vegetation density along the riverbanks, and temporal water-surface conditions were not recorded as explicit model inputs during the measurement campaign. Their possible influence on model transferability is therefore considered a limitation and is discussed in Section 4.

Figure 2 illustrates the field measurement data-collection process. The dataset consisted of 9,761 measurement points collected in 2023 along a 13.75 km measurement route on the

Table 1: List of equipment for data collection

Category	Equipment
Hardware	Laptop, Smartphone, GPS receiver
Software	TEMS Investigation
Supporting tools	Cable Data and USB Hub

Musi River. Measurements were conducted at 1800 MHz and 2100 MHz on LTE telecommunication networks. In this study, the transmitting BTS operated at an average transmitter power of 43 dBm, which was used in the measurement configuration for path-loss modeling. The transmitter height was calculated as the sum of the transmitting antenna height and the contour elevation at the transmitter location. Similarly, the receiver height was calculated as the sum of the receiving handset height and the contour elevation at the receiver location.

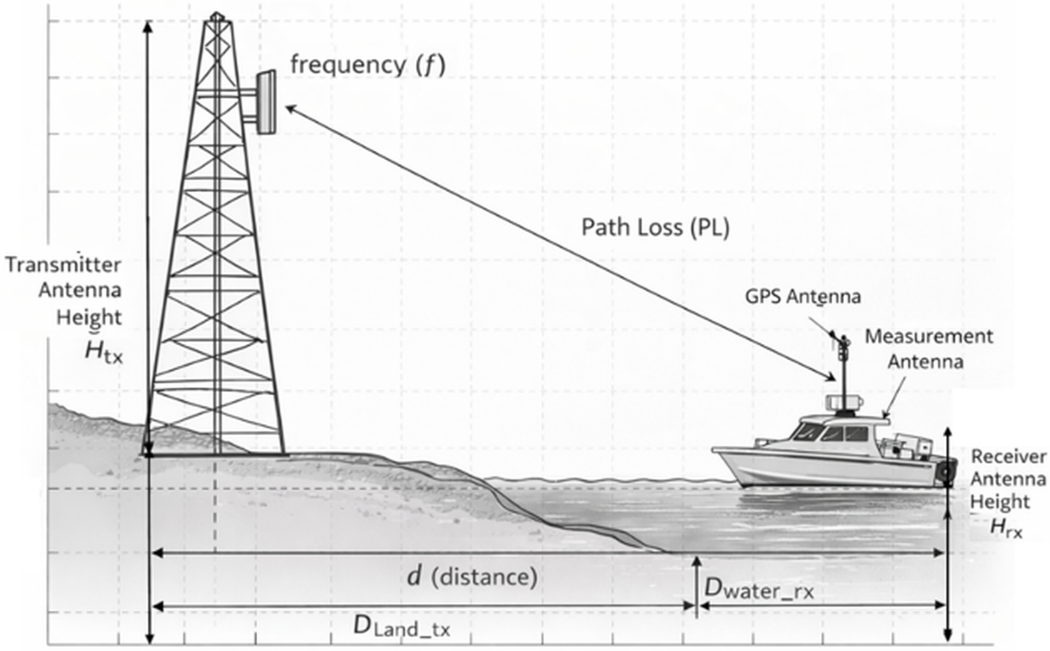


Figure 2: Field measurement data-collection process.

2.2 Data and Feature Engineering

The dataset collected from the Musi River included environmental and system parameters that influence radio-signal propagation. Based on physical relevance to propagation behavior, each observation included the following features in Table 2. Before model devel-

opment, all features were normalized to the range of -1 to 1 to improve numerical stability and reduce feature-scale bias during training.

Table 2: Definition of propagation model input parameters

Parameter	Symbol	Unit
Transmitter–receiver separation distance	d	m
Transmitter distance to nearest shoreline	$d_{\mathrm{land,tx}}$	m
Receiver distance to water surface	$d_{\mathrm{water,rx}}$	m
Carrier frequency	f	MHz
Transmitter antenna height	h_{tx}	m
Receiver antenna height	h_{rx}	m

2.3 Model Design and Implementation

This stage involved constructing, training, and optimizing the machine-learning and hybrid ensemble models used for path-loss prediction. RF and XGBoost were employed as base learners, while weighted blending and stacking were implemented as alternative ensemble strategies under the same experimental conditions.

2.3.1 Base Learners

The main base learners in the hybrid framework included several machine-learning algorithms with complementary learning characteristics. We selected Random Forest (RF) as an ensemble method based on multiple decision trees, providing robustness to noisy data and reducing the risk of overfitting by aggregating individual tree predictions. Meanwhile, XGBoost was incorporated as a gradient-boosting algorithm because it can capture complex nonlinear relationships and interactions among input features, making it effective for modeling highly variable propagation characteristics.

2.3.2 Stacking Hybrid Ensemble

In the stacking architecture, RF and XGBoost were employed as base learners, while SVR, Linear Regression, Ridge Regression, and Lasso Regression were evaluated as alternative meta-learners. Each meta-learner received out-of-fold predictions from the two base models as input, following the common practice of training the second-level learner on predictions that were not used to fit the base models [31, 33]. Their performance was compared using the same validation procedure and evaluation metrics. Based on the comparative results, SVR was selected as the final meta-learner because it achieved the lowest prediction errors while maintaining comparable R^2 performance.

2.3.3 Weighted Blending Hybrid Ensemble

The weighted blending approach integrates the prediction outputs of multiple machine learning models by assigning optimized weights to each base learner. In this framework, each model contributes to the final prediction based on its predictive strength, allowing the ensemble to leverage the complementary strengths of heterogeneous learners [32, 34].

The base learners, XGBoost and Random Forest (RF), were trained independently using GridSearchCV to determine optimal hyperparameter configurations. For XGBoost, the optimization process considered parameters such as learning rate, maximum tree depth, and the number of estimators, while the RF model was optimized based on parameters including the number of trees, tree depth, and feature selection strategy. We selected the optimal configurations based on the lowest validation error during cross-validation.

After optimizing each model, the prediction outputs from each base learner were combined using an adaptive weighted blending strategy. The final prediction of the ensemble model was formulated as:

$$\hat{y}_{\text{ensemble}} = w_1 \hat{y}_{\text{XGB}} + w_2 \hat{y}_{\text{RF}} \tag{1}$$

where $\hat{y}_{\text{ensemble}}$ represents the final path loss prediction produced by the hybrid ensemble model, $\hat{y}_{\text{XGB}}$ and $\hat{y}_{\text{RF}}$ denote the predictions generated by the XGBoost and Random Forest models, respectively, and w_1 and w_2 are the corresponding model weights. The weights satisfy the following constraints:

$$w_1 + w_2 = 1 \tag{2}$$

where $w_i \geq 0$.

'We determined the optimal weight combination empirically by minimizing prediction error on the validation dataset. The weight configuration that achieved the lowest error was selected as the final ensemble model for path loss prediction.

2.4 Model Training and Hyperparameter Optimization

Hyperparameter optimization was performed using `GridSearchCV` to identify suitable parameter configurations for each machine-learning model. The procedure systematically evaluated predefined hyperparameter combinations and selected the configuration that achieved the best cross-validation performance. The same optimization procedure was applied consistently across the individual learning algorithms to reduce parameter-selection bias and provide a fair basis for model comparison. The optimized hyperparameters for each model are shown in Table 3.

Table 3: The optimized hyperparameters for each model

Model	Hyperparameters	Parameter Search Range	Optimal Value
RF	n_estimators	100, 200, 300, 400, 500	300
	max_depth	10, 15, 20, 25, 30	20
	min_samples_split	2, 4, 6, 8, 10	6
XGBoost	learning_rate	0.01, 0.05, 0.10, 0.20, 0.30	0.1
	max_depth	3, 5, 7, 10	7
	n_estimators	100, 200, 300, 400	400
SVR	Kernel function	RBF	RBF
	Regularization parameter (C)	1, 10, 25, 50, 75, 100	10
	Epsilon	0.01, 0.05, 0.10, 0.15, 0.20	0.05

The dataset was evaluated using a randomized 10-fold cross-validation strategy. We shuffled the measurements before fold assignment so each validation fold represented a

mixture of propagation conditions observed along the measurement route. This strategy was intentionally retained because geographically contiguous blocks along the river may exhibit substantially different propagation characteristics due to variations in shoreline geometry, land–water configuration, surrounding structures, and antenna geometry. Using contiguous spatial blocks could therefore introduce strong domain differences between folds and confound model validation with site-specific environmental variation. The objective of the present validation was to assess predictive performance within the distribution of propagation conditions represented by the Musi River dataset rather than to estimate transferability to geographically unseen river sections. Nevertheless, because neighboring drive-test samples may remain spatially correlated, interpret the reported cross-validation performance as an estimate of within-dataset generalization rather than strict spatial extrapolation.

Out-of-fold predictions were generated using the same randomized 10-fold partitioning scheme for the stacking ensemble. Each sample was predicted only by base learners that were not trained on that sample, preventing direct target leakage into the meta-learner. However, this procedure does not eliminate spatial dependence among geographically adjacent observations, which is acknowledged as a limitation of the present validation design.

2.5 Empirical Propagation Model Benchmark

To establish conventional propagation baselines, four empirical path-loss models were evaluated: Okumura–Hata, ECC-33, COST-231 Hata, and Ericsson. These models were selected because they represent widely used empirical formulations with different frequency- and environment-dependent characteristics. For each measurement point, path loss was calculated using the corresponding operating frequency, transmitter–receiver distance, and antenna heights, and the resulting estimates were compared with the measured path loss values in Table 4.

Table 4: Performance of individual machine-learning models

Empirical Model	**Main Inputs**	**Nominal Frequency Range**	**Environmental Formulation**
Okumura–Hata	$d, f, H_{\mathrm{tx}}, H_{\mathrm{rx}}$	150–1500 MHz	Urban/suburban/open
COST-231 Hata	$d, f, H_{\mathrm{tx}}, H_{\mathrm{rx}}$	1500–2000 MHz	Urban/suburban
ECC-33	$d, f, H_{\mathrm{tx}}, H_{\mathrm{rx}}$	Higher cellular bands	Urban/medium city formulation
Ericsson	$d, f, H_{\mathrm{tx}}, H_{\mathrm{rx}}$	Configurable empirical model	Adjustable environment coefficients

Because part of the measurement dataset was collected at 2100 MHz, which is slightly outside the nominal frequency range of some classical empirical models, their results are treated as comparative empirical baselines rather than as strictly in-range predictions.

2.6 Performance Metrics

The performance evaluation stage was conducted to assess and compare the predictive accuracy of the empirical, individual machine-learning, and hybrid ensemble models under

the same measurement conditions. Model performance was quantified using standard regression metrics to provide a consistent basis for determining the most accurate path-loss prediction approach:

$$\mathrm{MAE} = \frac{1}{N}\sum_{j=1}^{n}|y_j - \hat{y}_j|, \tag{3}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}, \tag{4}$$

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - \hat{y}_j)^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2}, \tag{5}$$

where y_i denotes the measured path loss value, $\hat{y}_i$ represents the predicted path loss value, $\bar{y}$ is the mean of the observed measured path loss values, and n is the total number of observations.

3 Results

3.1 Feature Correlation Assessment

A feature correlation assessment was conducted to examine the linear relationships among the input variables and their associations with the measured path-loss values. Pearson correlation coefficients were used to identify the strength and direction of these relationships and to provide an initial understanding of potential interdependencies among the predictors before model development.

Figure 3 shows correlations between the input features and measured path-loss values. Frequency has the strongest positive correlation with path loss ($r = 0.76$), followed by transmitter–receiver distance ($r = 0.40$). Several spatial variables are also strongly correlated, particularly d with $d_{\mathrm{land,rx}}$ ($r = 0.78$) and $d_{\mathrm{land,tx}}$ ($r = 0.72$). These results indicate that the predictors contain both complementary and partially redundant information. However, because correlation reflects only pairwise linear relationships, we used a separate feature-importance analysis to evaluate each variable's relative contribution to the prediction model.

3.2 Feature Importance Analysis

A feature-importance analysis was conducted to quantify the relative predictive contribution of the six input variables used in the model. We performed this analysis as a post-hoc interpretability assessment rather than a feature-selection procedure, because we retained all predictors based on their physical relevance to radio propagation and riverine geometry. The resulting normalized importance scores are presented in Table 5.

The normalized feature-importance analysis shows that frequency was the most influential predictor, accounting for 44.13% of the total importance. Transmitter–receiver distance ranked second at 18.11%, followed closely by $d_{\mathrm{land,tx}}$ at 17.34%. Receiver antenna height contributed 10.39%, while $d_{\mathrm{land,rx}}$ and h_{tx} contributed 5.24% and 4.78%, respectively. These results indicate that both conventional propagation variables and river-related

spatial parameters contributed to the prediction process. In particular, the relatively high importance of $d_{\mathrm{land,tx}}$ supports including land–water spatial information in the proposed framework. Although $d_{\mathrm{land,rx}}$ and h_{rx} exhibited lower relative importance, their non-zero contributions indicate that they still provided complementary predictive information and were therefore retained in the model.

Table 5: List of equipment for data collection

Rank	Feature	Normalized Importance (%)
1	frequency	44.12512543
2	d	18.11475883
3	$d_{\mathrm{land,tx}}$	17.34322049
4	h_{rx}	10.39414838
5	$d_{\mathrm{land,rx}}$	5.242361412
6	h_{tx}	4.78038545

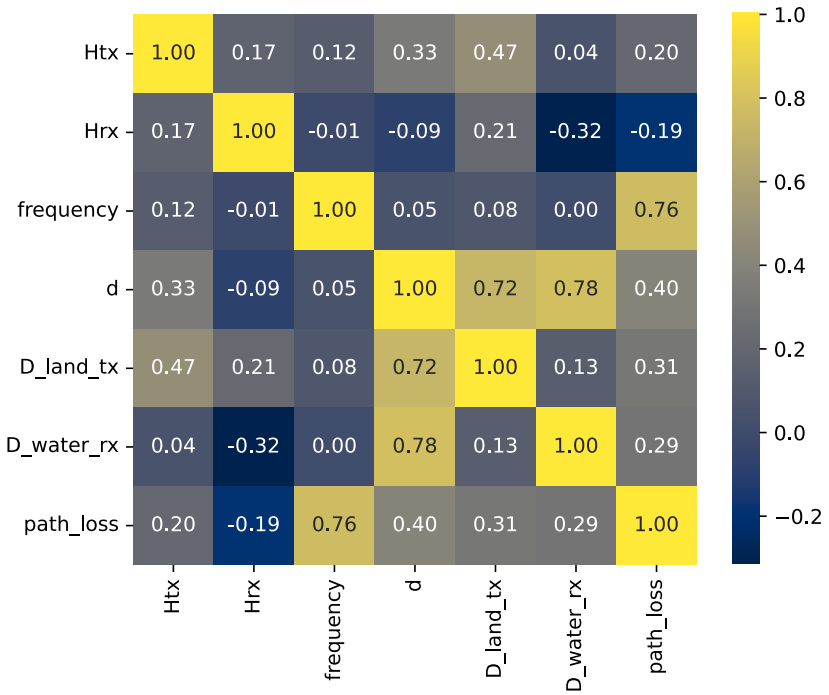


Figure 3: Correlation matrix of input features and measured path loss.

3.3 Performance of Empirical Propagation Models

To establish conventional propagation baselines, four empirical models were evaluated using the same measurement dataset, namely Okumura–Hata, ECC-33, COST-231 Hata, and

Ericsson. We assessed each model's performance using mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) to provide a quantitative basis for comparison with the machine-learning and hybrid ensemble approaches in Table 6.

Table 6: Performance of empirical propagation models

Empirical Model	**MAE**	**RMSE**	R^2
Okumura–Hata	10.9479	13.2097	0.1683
ECC-33	23.7689	26.7844	-2.4193
COST-231	10.7504	13.1797	0.1721
Ericsson	16.077	19.4311	-0.7996

Among the evaluated empirical models, COST-231 performed best, followed closely by Okumura-Hata. Ericsson produced higher prediction errors, while ECC-33 showed the weakest agreement with the measured data, as reflected by its negative R^2 value. Overall, the results suggest that conventional empirical models have a limited ability to accurately represent the propagation characteristics of the Musi River environment.

3.4 Predictive Accuracy of Base Machine-Learning Models

The predictive performance of the individual base learners was first evaluated to establish a baseline before applying the hybrid ensemble strategies. Random Forest and XGBoost were assessed using the same dataset, validation procedure, and performance metrics to ensure a consistent comparison. Their predictive accuracy is shown in Table 7.

Table 7: Performance of individual machine-learning models

Single Model	**Performance Evaluation**		
	MAE	**RMSE**	R^2
Random Forest	2.8524	3.8004	0.9312
XGBoost	2.9811	3.8595	0.929

Table 7 compares the predictive performance of the individual machine-learning models. Random Forest performed slightly better than XGBoost, with a lower MAE of 2.8524 dB and RMSE of 3.8004 dB, and a higher R^2 of 0.9312. XGBoost produced an MAE of 2.9811 dB, RMSE of 3.8595 dB, and R^2 of 0.9290. The relatively small differences between the two models indicate that both base learners provide strong predictive performance and are suitable for integration within the proposed hybrid ensemble framework.

3.5 Weighted Hybrid Prediction Model

The weighted hybrid prediction model combines Random Forest and XGBoost predictions using weighted averaging. Each base learner's contribution is controlled by its assigned weight, and the model selects the final weighted combination that minimizes prediction error.

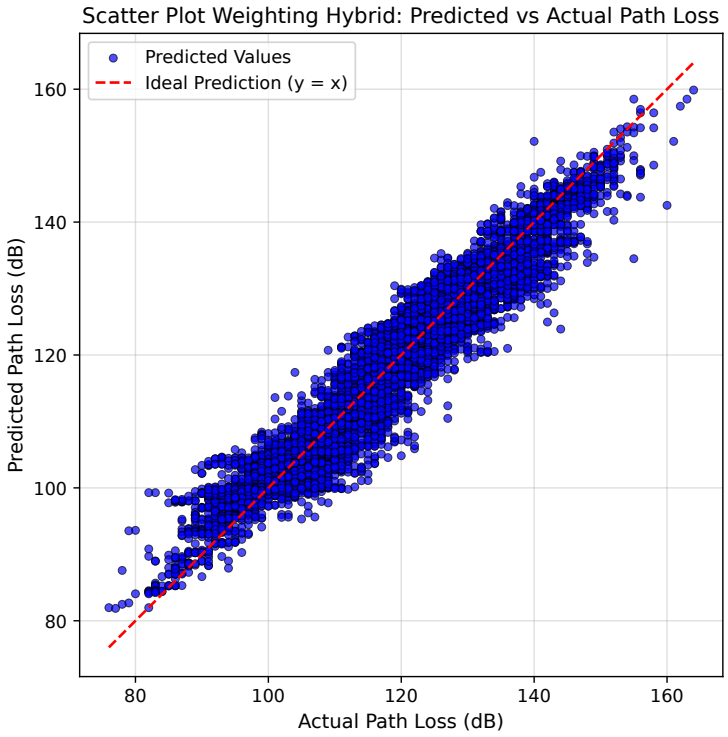


Figure 4: Measured versus predicted path loss for the weighted hybrid model.

Table 8: Performance of weighted blending hybrid models

Weighting Hybrid	Performance Evaluation		
	MAE	**RMSE**	R^2
$0.9\hat{y}_{\mathrm{RF}} + 0.1\hat{y}_{\mathrm{XGB}}$	2.84380	3.77000	0.9320
$0.8\hat{y}_{\mathrm{RF}} + 0.2\hat{y}_{\mathrm{XGB}}$	2.84070	3.74740	0.9330
$0.7\hat{y}_{\mathrm{RF}} + 0.3\hat{y}_{\mathrm{XGB}}$	2.84240	3.73300	0.9340
$0.6\hat{y}_{\mathrm{RF}} + 0.4\hat{y}_{\mathrm{XGB}}$	2.84950	3.72670	0.9340
$0.5\hat{y}_{\mathrm{RF}} + 0.5\hat{y}_{\mathrm{XGB}}$	2.86080	3.72870	0.9340
$0.4\hat{y}_{\mathrm{RF}} + 0.6\hat{y}_{\mathrm{XGB}}$	2.87610	3.73880	0.9330
$0.3\hat{y}_{\mathrm{RF}} + 0.7\hat{y}_{\mathrm{XGB}}$	2.89570	3.75710	0.9330
$0.2\hat{y}_{\mathrm{RF}} + 0.8\hat{y}_{\mathrm{XGB}}$	2.91950	3.78350	0.9320
$0.1\hat{y}_{\mathrm{RF}} + 0.9\hat{y}_{\mathrm{XGB}}$	2.94820	3.81760	0.9310

The results in Table 8 show that the weighting configuration affects the hybrid model's predictive performance. Among the evaluated combinations, the configuration producing an MAE of 2.8407 dB achieved the lowest average absolute error, while the lowest RMSE of 3.7267 dB was obtained by another weighting combination. The highest R^2 value reached 0.9340 across several configurations. Overall, the relatively small differences among the tested weights indicate that the weighted hybrid model remains stable across different RF–XGBoost combinations. However, the selected configuration should reflect the primary evaluation criterion.

Figure 4 shows the agreement between the measured and predicted path-loss values from the weighted hybrid model. Most data points lie close to the ideal $y = x$ reference line, indicating that the model's predictions generally match the measured values. The relatively small dispersion around the reference line also reflects the weighted blending approach's strong predictive capability.

3.6 Stacking Hybrid Ensemble

To justify the meta-learner selection in the stacking architecture, we compared SVR with Lasso Regression, Linear Regression, and Ridge Regression using the same RF and XGBoost base predictions. The comparative performance of these meta-learners is presented in Table 9.

Table 9: Performance of hybrid models

Hybrid Model	Performance Evaluation		
	MAE	**RMSE**	R^2
Stacking Hybrid–SVR	2.6213	3.6885	0.9352
Stacking Hybrid–Lasso Regression	2.6352	3.6888	0.9355
Stacking Hybrid–Linear Regression	2.6356	3.6896	0.9355
Stacking Hybrid–Ridge Regression	2.6357	3.6898	0.9355

The results show that the SVR-based stacking model achieved the lowest MAE and RMSE among the evaluated meta-learners, at 2.6213 dB and 3.6885 dB, respectively. Al-

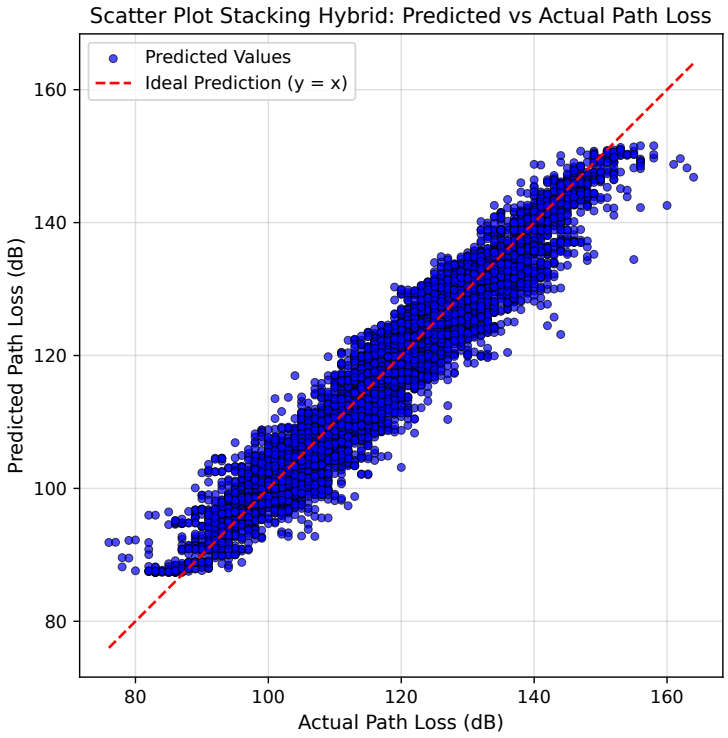


Figure 5: Scatter plot of measured versus predicted path-loss values for the stacking hybrid SVR.

though Lasso, Linear Regression, and Ridge Regression produced slightly higher R^2 values of 0.9355, the differences were marginal. Therefore, SVR was retained as the final meta-learner because it provided the lowest prediction error while maintaining comparable goodness-of-fit.

The scatter plot in Figure 5 shows that most predicted values are closely distributed around the ideal $y = x$ reference line, indicating strong agreement between the measured and predicted path-loss values. Although several deviations are observed at the lower and higher path-loss ranges, the overall distribution demonstrates the strong predictive capability of the stacking hybrid SVR model.

3.7 Comparative Performance of Hybrid Models

A direct comparison was conducted to assess the relative predictive performance of the two proposed hybrid ensemble models. Their performance was evaluated using MAE, RMSE, and R^2, and the results are summarized in Table 10. The stacking hybrid with SVR outperformed the weighting hybrid across all evaluation metrics, achieving lower MAE and RMSE values and a slightly higher R^2. This result indicates that the stacking strategy provides a more effective combination of the base-learner predictions than the fixed weighted-blending approach.

Table 10: Performance of hybrid models

Hybrid Model	**Performance Evaluation**		
	MAE	**RMSE**	R^2
Weighting Hybrid	2.8407	3.7474	0.9331
Stacking Hybrid (SVR)	2.6213	3.6885	0.9352

Figure 6 illustrates the residual distribution of the stacking and weighted-blending models across the measurement samples. The residuals cluster predominantly around the zero-reference line, indicating that both hybrid models generally produced predictions close to the measured path-loss values without strong systematic bias. The stacking model shows a slightly more concentrated residual distribution around zero than the weighted-blending model, consistent with the quantitative results in Table 4. The stacking hybrid achieved lower MAE and RMSE values of 2.6213 dB and 3.6885 dB, respectively, together with a higher R^2 value of 0.9352. In comparison, the weighted-blending hybrid produced an MAE of 2.8407 dB, an RMSE of 3.7474 dB, and an R^2 of 0.9331. These results indicate that the stacking architecture integrates the RF and XGBoost predictions slightly more effectively, although both ensemble approaches show comparable overall predictive performance.

3.8 Comprehensive Comparison of Empirical, Base and Hybrid Models

A comprehensive performance comparison was conducted among the empirical propagation models, individual machine-learning models, and hybrid ensemble approaches to evaluate their relative prediction capability under the same measurement conditions. The corresponding MAE, RMSE, and R^2 values are presented in Table 11.



Figure 6: Residual distribution of the stacking and weighted-blending models across the measurement samples.

Table 11: Comparative performance of empirical, machine-learning, and hybrid models

Model	**Model Category**	**MAE**	**RMSE**	R^2
Okumura–Hata Model	Empirical	10.9479	13.2097	0.1683
ECC-33 Model	Empirical	23.7689	26.7844	-2.4193
COST231–Hata Model	Empirical	10.7504	13.1797	0.1721
Ericsson Model	Empirical	16.077	19.4311	-0.7996
Random Forest	ML	2.8524	3.8004	0.9312
XGBoost	ML	2.9811	3.8595	0.929
Weighted Blending	Hybrid	2.8407	3.7474	0.9331
Stacking	Hybrid	2.6213	3.6885	0.9352

The overall comparison shows a clear performance advantage for the machine-learning and hybrid approaches over conventional empirical propagation models. Among the empirical models, COST231-Hata performed best, with an MAE of 10.7504 dB, an RMSE of 13.1797 dB, and an R^2 of 0.1721, whereas ECC-33 showed the weakest agreement with the measured data. The individual machine-learning models substantially reduced the prediction error, with Random Forest performing slightly better than XGBoost. Further improvement was obtained through the hybrid ensemble strategies. The weighted-blending model achieved an MAE of 2.8407 dB and an RMSE of 3.7474 dB, while the stacking model provided the best overall performance, with the lowest MAE of 2.6213 dB, the lowest RMSE of 3.6885 dB, and the highest R^2 of 0.9352. These results indicate that the stacking architecture offers the strongest predictive performance among the evaluated models for the Musi River propagation dataset.

4 Discussion

Compared with conventional empirical propagation models, data-driven approaches show a substantial performance advantage. Among the empirical models, COST-231 performed best, with an RMSE of 13.1797 dB and an MAE of 10.7504 dB. In contrast, the stacking hybrid achieved a substantially lower RMSE of 3.6885 dB and an MAE of 2.6213 dB. This corresponds to an RMSE reduction of approximately 72.0% and an MAE reduction of approximately 75.6% relative to the best-performing empirical baseline. The stacking model also achieved an R^2 of 0.9352, substantially higher than the R^2 of 0.1721 obtained by COST-231. These results indicate that the data-driven models, particularly the stacking hybrid, better represented the propagation characteristics captured in the Musi River measurement dataset than the evaluated empirical formulations.

Among the machine-learning approaches, both hybrid ensemble models outperformed the individual benchmark models, although the improvement over Random Forest remained relatively modest. The stacking-based hybrid model delivered the highest overall prediction accuracy, obtaining an MAE of 2.6213 dB, an RMSE of 3.6885 dB, and an R^2 of 0.9352. This outcome shows that the stacking strategy provided a more effective way to combine the complementary prediction strengths of RF and XGBoost than the fixed weighted-blending scheme. In particular, the SVR meta-learner provided an adaptive nonlinear combination of the base-model outputs, resulting in lower prediction errors while maintaining strong agreement with the measured path-loss values.

This performance difference may also reflect the predictor information available to the data-driven framework. Conventional empirical models primarily describe path loss using predefined relationships among distance, frequency, antenna heights, and environment-dependent correction factors. In contrast, the proposed framework additionally incorporates river-related spatial features, including $d_{\mathrm{land,tx}}$ and $d_{\mathrm{land,rx}}$, which provide additional information regarding the land–water propagation geometry. These features may help the learning algorithms represent localized propagation effects associated with water reflections, shoreline geometry, and spatial variations along the river environment.

The feature-importance results further support the selected predictor set. Frequency, transmitter–receiver distance, and $d_{\mathrm{land,tx}}$ were the three most influential variables, indicating that both conventional propagation parameters and river-specific spatial information contributed to the model's predictions. The relatively high importance of $d_{\mathrm{land,tx}}$ suggests that the transmitter's position relative to the land boundary provides useful information beyond distance and frequency alone. At the same time, the lower but non-zero importance of $d_{\mathrm{land,rx}}$ and h_{tx} indicates that these features still contribute complementary information and should not be considered irrelevant solely because of their lower ranking. However, the feature-importance results should be interpreted as relative predictive contributions rather than as evidence of causal effects. Because several predictors are correlated, part of their predictive information may be shared across variables. Moreover, no feature-ablation experiment was conducted; therefore, the present analysis cannot conclusively isolate the individual effect of each predictor on overall model performance.

From a telecommunications perspective, the remaining variation in prediction error may be associated with local propagation characteristics that are not explicitly represented by the current predictor set. Measurements obtained near the land-water boundary may be influenced by surrounding built-up structures. Building height can alter line-of-sight and non-line-of-sight conditions and introduce additional shadowing, diffraction, and block-

age, while building density can increase multipath propagation through reflections and scattering from multiple surfaces. Consequently, geographically close measurement points with similar transmitter–receiver distances or land water boundary distances may still experience different propagation conditions because of differences in the surrounding built environment. In the present study, parameters such as building height, building density, obstruction profile, and shoreline structural characteristics were not explicitly included in the model. These unobserved environmental factors may therefore partially explain the localized fluctuations that remain in the prediction errors. From a network-planning perspective, these residual variations matter because localized underestimation or overestimation of path loss can affect coverage-margin estimation, link-budget calculation, and infrastructure-placement decisions. However, because these environmental characteristics were not directly measured, the present dataset cannot quantify their influence, and they should be interpreted as a plausible propagation-related limitation rather than a confirmed causal explanation.

The contribution of this study should therefore be interpreted as application-oriented and comparative rather than as the introduction of a new propagation dataset or a fundamentally new machine-learning algorithm. The riverine measurement data provides the experimental basis. At the same time, the principal methodological contribution is evaluating how established base learners can be integrated through weighted blending and stacking, and quantitatively comparing these approaches with empirical propagation models and individual machine-learning baselines. The additional feature-importance analysis further clarifies the role of conventional and river-related predictors in the final prediction framework.

A limitation of the present validation design is the spatial dependence inherent in sequential drive-test measurements. Although randomized 10-fold cross-validation was retained to represent the heterogeneous propagation conditions observed along the complete measurement corridor, this approach does not fully eliminate correlation among geographically adjacent samples. Therefore, the reported performance should be interpreted as within-dataset generalization rather than strict spatial extrapolation to unseen river sections.

This study has several additional limitations that should be considered when interpreting the results. First, the measurement campaign was conducted only along the Musi River; therefore, the reported model performance primarily represents the propagation characteristics of this specific river environment and cannot yet be generalized to other riverine settings. Second, the measurements were limited to the LTE frequency bands investigated in this study, namely 1800 MHz and 2100 MHz, so the performance of the proposed framework at other frequency bands remains to be evaluated. Third, environmental variables such as water level were not recorded during the measurement campaign. Variations in water level may alter the effective land–water geometry, shoreline conditions, and reflection characteristics along the propagation path. In addition, vegetation characteristics, building height, and building density along the riverbanks were not explicitly incorporated as model features. These environmental factors may influence LOS/NLOS conditions, shadowing, diffraction, scattering, blockage, and multipath propagation, particularly near built-up or vegetated shoreline areas. Consequently, some localized prediction errors may reflect propagation effects that are not fully represented by the current predictor set.

The proposed models provide a strong foundation for path-loss prediction in riverine environments, having been developed and evaluated using measurement data from a rep-

resentative river scenario and selected LTE frequency bands. Their applicability can be further extended by validating them across additional river environments and frequency conditions. River propagation conditions may vary according to environmental characteristics that were not explicitly included as predictors in the present dataset. For example, river width can modify the relative contribution of over-water propagation, shoreline reflections, and the geometric relationship between the transmitter and receiver. Current velocity may indirectly influence propagation by affecting water-surface roughness and temporal surface conditions, which can alter the characteristics of reflected components. Vegetation along the riverbanks may also introduce additional attenuation, shadowing, and scattering, particularly when the propagation path is close to or partially obstructed by vegetated areas. Other site-specific characteristics, such as riverbank geometry, surrounding structures, water level, and surface conditions, may further contribute to differences in propagation behavior between river environments. Because these environmental variables were not explicitly incorporated into the current feature set, the reported performance primarily represents the measurement conditions of the Musi River dataset. Consequently, external validation using measurements from rivers with different widths, flow conditions, shoreline vegetation, and surrounding morphologies is required before the model can be considered transferable across riverine environments. Future studies should therefore incorporate these environmental descriptors as additional model features and evaluate their contribution to prediction accuracy and cross-river generalization.

5 Conclusion

This study applied a hybrid ensemble learning framework to predict path loss in a riverine environment using measurement data collected along the Musi River. The study's main contribution is the systematic evaluation of hybrid ensemble learning for riverine path-loss prediction, rather than the introduction of a new dataset or learning algorithm. By comparing conventional empirical propagation models, individual machine-learning approaches, weighted blending, and stacking under the same measurement conditions, the study demonstrates the practical value of ensemble learning for the Musi River propagation scenario. The results showed that the hybrid ensemble models substantially outperformed the conventional empirical propagation models and improved upon the individual machine-learning models. Among the empirical baselines, ECC-33 achieved the lowest RMSE of 13.1797 dB, whereas the stacking hybrid reduced the RMSE to 3.6915 dB and achieved an MAE of 2.7802 dB and an R^2 of 0.9350. This corresponds to an approximately 72.0% reduction in RMSE relative to the best-performing empirical baseline. Among the evaluated hybrid approaches, stacking also outperformed weighted blending, indicating that the SVR-based meta-learning mechanism was more effective than a fixed weighted-average scheme at integrating complementary RF and XGBoost predictions. These findings indicate that data-driven ensemble learning can provide more accurate path-loss prediction for the riverine propagation conditions represented by the Musi River dataset and can support more reliable wireless-network planning in over-water communication environments. Future work should extend the proposed framework to independent measurement datasets collected from other river environments in Indonesia to evaluate cross-river generalization. Additional environmental descriptors, including water level, shoreline vegetation, building height and density, river width, and surface conditions, should also be

incorporated to represent localized propagation effects better. Furthermore, the proposed ensemble framework can be compared with more advanced deep-learning architectures, such as Long Short-Term Memory (LSTM) networks and Transformer-based models, to assess whether these approaches can further improve modeling of complex spatial and sequential propagation characteristics.

Acknowledgments

The authors acknowledge the financial support provided by Universitas Muhammadiyah Palembang, which made this research possible.

References

[1] Y. H. Santana, R. M. Alonso, G. G. Nieto, L. Martens, W. Joseph, and D. Plets, "5G mmWave network planning using machine learning for path loss estimation," *IEEE Open Journal of the Communications Society*, vol. 5, pp. 3451–3467, 2024.

[2] A. Gupta, J. Du, D. Chizhik, R. A. Valenzuela, and M. Sellathurai, "Machine learning-based urban canyon path loss prediction using 28 GHz manhattan measurements," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 4096–4111, 2022.

[3] J. Qian, Y. Wu, A. Saleem, and G. Zheng, "Path loss model for 3.5 GHz and 5.6 GHz bands in cascaded tunnel environments," *Sensors*, vol. 22, no. 12, p. 4524, 2022.

[4] A. Kaya, B. De Beelde, W. Joseph, M. Weyn, and R. Berkvens, "Geodesic path model for indoor propagation loss prediction of narrowband channels," *Sensors*, vol. 22, no. 13, p. 4903, 2022.

[5] O. J. Famoriji, "Path loss prediction in tropical regions using machine learning techniques: A case study," *Electronics*, vol. 11, no. 17, p. 2711, 2022.

[6] C. Nguyen and A. A. Cheema, "A deep neural network-based multi-frequency path loss prediction model from 0.8 GHz to 70 GHz," *Sensors*, vol. 21, no. 15, p. 5100, 2021.

[7] B. Kwon and H. Son, "Accurate path loss prediction using a neural network ensemble method," *Sensors*, vol. 24, no. 1, p. 304, 2024.

[8] M. K. Elmezughi, O. Salih, T. J. Afullo, and K. J. Duffy, "Path loss modeling based on neural networks and ensemble method for future wireless networks," *Heliyon*, vol. 9, no. 9, p. e19685, 2023.

[9] M. A. Samad, S. W. Choi, C. S. Kim, and K. Choi, "Wave propagation modeling techniques in tunnel environments: A survey," *IEEE Access*, vol. 11, pp. 2199–2225, 2023.

[10] M. Sousa, A. Alves, P. Vieira, M. P. Queluz, and A. Rodrigues, "Analysis and optimization of 5G coverage predictions using a beamforming antenna model and real drive test measurements," *IEEE Access*, vol. 9, pp. 101787–101808, 2021.

[11] K. Qiu, S. Bakirtzis, H. Song, J. Zhang, and I. Wassell, "Pseudo ray-tracing: Deep learning assisted outdoor mmWave path loss prediction," *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1699–1702, 2022.

[12] I. Iliev, Y. Velchev, P. Z. Petkov, B. Bonev, G. Iliev, and I. Nachev, "A machine learning approach for path loss prediction using combination of regression and classification models," *Sensors*, vol. 24, p. 5855, 2024.

[13] G. Pau and L. B. Le, "LoRaWAN path loss measurements in an urban scenario including environmental effects," *Data*, vol. 8, no. 1, p. 4, 2023.

[14] M. O. Ojo, I. Viola, S. Miretti, E. Martignani, S. Giordano, and M. Baratta, "A deep learning approach for accurate path loss prediction in LoRaWAN livestock monitoring," *Sensors*, vol. 24, no. 10, p. 2991, 2024.

[15] A. I. Griva *et al.*, "LoRa-based IoT network assessment in rural and urban scenarios," *Sensors*, vol. 23, no. 3, p. 1695, 2023.

[16] S. Phaiboon and P. Phokharatkul, "Multi-boundary empirical path loss model for 433 MHz WSN in agriculture areas using fuzzy linear regression," *Sensors*, vol. 23, no. 7, p. 3525, 2023.

[17] H. Vo, V. H. L. Nguyen, V. L. Tran, F. Ferrero, F. Y. Lee, and M. H. Tsai, "Advanced path loss model for distance estimation using LoRaWAN network's received signal strength indicator (RSSI)," *IEEE Access*, vol. 12, pp. 83205–83216, 2024.

[18] M. Bouzidi, M. Mohamed, Y. Dalveren, A. Moldsvor, F. A. Cheikh, and M. Derawi, "Propagation measurements for IQRF network in an urban environment," *Sensors*, vol. 22, no. 18, p. 7012, 2022.

[19] M. G. Gaitan *et al.*, "Modeling LoRa communications in estuaries for IoT environmental monitoring systems," *IEEE Sensors Journal*, vol. 22, no. 21, pp. 21312–21325, 2022.

[20] S. Shen, W. Zhang, H. Zhang, Q. Ren, X. Zhang, and Y. Li, "An accurate maritime radio propagation loss prediction approach employing neural networks," *Remote Sensing*, vol. 14, p. 4753, 2022.

[21] A. Lindenbergs *et al.*, "Seamless 5G multi-hop connectivity architecture and trials for maritime applications," *Sensors*, vol. 23, no. 9, p. 4203, 2023.

[22] S. Duangsuwan, "Comparison of path loss prediction models for UAV and IoT air-to-ground communication system in rural precision farming environment," *Journal of Communications*, vol. 16, no. 2, pp. 60–66, 2021.

[23] H. Li *et al.*, "Air-to-ground path loss model at 3.6 GHz under agricultural scenarios based on measurements and artificial neural networks," *Drones*, vol. 7, no. 12, p. 701, 2023.

[24] V. C. Ebhota and V. M. Srivastava, "Effect of architectural composition of MLP ANN in neural network learning for signal power loss prediction," *Journal of Communications*, vol. 16, no. 1, pp. 20–29, 2021.

[25] C. Huang *et al.*, "Artificial intelligence enabled radio propagation for communications—part I: Channel characterization and antenna-channel optimization," *IEEE Transactions on Antennas and Propagation*, vol. 70, no. 6, pp. 3939–3954, 2022.

[26] N. Tarhuni, I. Al Saadi, H. M. Asif, M. Mesbah, O. Eldirdiry, and A. Hossen, "Machine-learning-based ground-level mobile network coverage prediction using UAV measurements," *Journal of Sensor and Actuator Networks*, vol. 12, no. 3, p. 44, 2023.

[27] L. Ho and P. Goethals, "Machine learning applications in river research: Trends, opportunities and challenges," *Methods in Ecology and Evolution*, vol. 13, no. 11, pp. 2603–2621, 2022.

[28] T. L. Carlson, L. A. P. LaBrie, J. S. Wesner, S. R. Chipps, A. A. Coulter, and B. J. Schall, "Receiver mount design, transmitter depth, and wind speed affect detection probability of acoustic telemetry transmitters in a missouri river tributary," *Animal Biotelemetry*, vol. 11, p. 6, 2023.

[29] E. Moradi and L. Miranda-Moreno, "A mixed ensemble learning and time-series methodology for category-specific vehicular energy and emissions modeling," *Sustainability*, vol. 14, no. 3, p. 1900, 2022.

[30] J. O. Afape *et al.*, "Improving millimetre-wave path loss estimation using automated hyperparameter-tuned stacking ensemble regression machine learning," *Results in Engineering*, vol. 22, p. 102289, 2024.

[31] J. Gu, S. Liu, Z. Zhou, S. R. Chalov, and Q. Zhuang, "A stacking ensemble learning model for monthly rainfall prediction in the Taihu basin, china," *Water*, vol. 14, no. 3, p. 492, 2022.

[32] M. Herrmann and W. Marzocchi, "Maximizing the forecasting skill of an ensemble model," *Geophysical Journal International*, vol. 234, no. 1, pp. 73–87, 2023.

[33] A. S. Mohammed, P. G. Asteris, M. Koopialipoor, D. E. Alexakis, M. E. Lemonis, and D. J. Armaghani, "Stacking ensemble tree models to predict energy performance in residential buildings," *Sustainability*, vol. 13, no. 15, p. 8298, 2021.

[34] A. Ray, T. Chakraborty, and D. Ghosh, "Optimized ensemble deep learning framework for scalable forecasting of dynamics containing extreme events," *Chaos*, vol. 31, no. 11, p. 113113, 2021.

[35] L. Peng, K. Yang, J. Wu, C. Wen, T. Peng, and X. Li, "Machine learning methods comparison for maritime wireless signal strength prediction," *Engineering Applications of Artificial Intelligence*, vol. 158, p. 111357, 2025.

[36] J. Ding, K. Yang, L. Qin, and B. Zheng, "Prediction model for maritime 5G signal strength based on ConvLSTM-PSO-XGBoost algorithm," *Journal of Marine Science and Engineering*, vol. 14, no. 4, p. 377, 2026.

[37] N. Sağır, E. Eyüboğlu, K. Kuzulugil, Z. Hasırcı Tuğcu, and İ. H. Çavdar, "Real maritime measurement-driven machine learning for path loss modeling in shore-to-ship wireless links," *Journal of Engineering Research*, 2026.

[38] B. Alfaresi, Z. Nawawi, and B. Y. Suprapto, "Development of path loss prediction model using feature selection-machine learning approach," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 10, pp. 349–358, 2022.

[39] B. Shu, W. Zhang, Y. Chen, J. Sun, and C.-X. Wang, "Path loss prediction in evaporation ducts based on deep neural network," *IEEE Antennas and Wireless Propagation Letters*, vol. 23, no. 2, pp. 798–802, 2024.

[40] S. A. Aldossari, "Predicting path loss of an indoor environment using artificial intelligence in the 28-GHz band," *Electronics*, vol. 12, no. 3, p. 497, 2023.