Data-Driven Path Loss Modeling Using Multilayer Perceptron Networks

1st Dong-Seok Lee
dept. Automotive Engineering
(Automotive–Computer Convergence)
Hanyang University
Seoul, Republic of Korea
dseok2805@hanyang.ac.kr

2nd Taesik Nam
dept. Electrical and Electronic Engineering
Yonsei University
Seoul, Republic of Korea
ts.nam@yonsei.ac.kr

3rd Han-Shin Jo*
dept. Automotive Engineering
Hanyang University
Seoul, Republic of Korea
hsjo23@hanyang.ac.kr
(Corresponding Author)

***Abstract*—This study compares ML/DL-based path loss prediction models using empirically measured data while accounting for regional nonlinear propagation characteristics. Multivariable Linear Regression (MLR), Support Vector Regression (SVR), and a Multi-Layer Perceptron (MLP) were trained using log-transformed representations of distance and frequency, and their generalization performance was evaluated in two regions (Area A and Area B). The experimental results revealed that in Area A, characterized by strong linearity, the predictive performance of the three models was generally comparable; however, in Area B, where nonlinearity was pronounced, both SVR and MLP exhibited higher R^2 values and lower RMSE compared to MLR. In particular, the MLP model was able to capture small nonlinear variations even in the mostly linear characteristics of Area A. In Area B, where propagation fluctuates more severely due to environmental factors, it achieved the highest prediction accuracy among the evaluated models. These findings highlight the potential of ML/DL-based nonlinear approaches to improve path loss prediction accuracy in diverse wireless channels.**

***Index Terms*—Path loss prediction, radio propagation, nonlinear regression, SVR, MLP, machine learning.**

I. Introduction

As confirmed by recent developments in International Telecommunication Union – Radiocommunication Sector (ITU-R) Study Group 3 (SG3) and ITU-R Question 236/3[1], international interest in radio wave propagation prediction using machine learning is rapidly increasing. The ITU-R has emphasized that data collection and processing methodologies, along with modeling techniques, need to be incorporated into future ITU-R reports, recommendations, and handbooks, while discussing the standardization of machine learning-based propagation prediction methods from various perspectives including physics-based modeling and explainability, domain adaptation, data augmentation, uncertainty quantification, model optimization, and validation based on real-world environments. This means that the importance of data-driven approaches in the field of radio propagation prediction is emerging at the international standard level.

Radio wave propagation prediction is an essential component in various fields, including cellular communications, satellite-terrestrial interference analysis, and wireless network design, with accurate path loss modeling serving as the cornerstone of wireless network performance prediction. Although conventional propagation models are based on physical factors such as frequency, distance, and terrain data, they are inherently limited in that a single model cannot adequately explain all situations due to the countless variables and irregularities that exist in real-world environments. In a previous study [2], a new path loss model based on the modified Hata model [3] was proposed for actual measured path loss data from two regions. Although the proposed model demonstrated high accuracy in Region A, where the measurement data exhibited strong linear characteristics, it showed limitations in accurately capturing subtle nonlinear variations particularly in distance ranges where the path loss increased rapidly and subsequently plateaued. In contrast, Region B presented pronounced nonlinearity and severe variability in the data due to complex environmental factors such as building distribution and scattering effects, making it difficult to apply linear modeling approaches. Consequently, when the nonlinearity of the propagation environment is strong in different regions, simple linear regression alone cannot adequately learn the complex variation patterns in the data, necessitating nonlinear regression techniques or flexible modeling approaches based on machine learning and deep learning.

In this study, predictive modeling of radio wave path loss was conducted by applying multiple linear regression (MLR), radial basis function (RBF) kernel-based support vector regression (SVR), and multilayer perceptron (MLP) based on the empirical data from [2]. The prediction accuracy of the three techniques was quantitatively compared using Coefficient of Determination (R^2) and Root Mean Squared Error (RMSE), and the limitations identified during the analysis process were summarized. Based on these findings, future directions for improving radio propagation prediction models are proposed, including data diversification and nonlinear model expansion.

II. Regression Method

A. Multivariable Linear Regression

Linear regression is a fundamental machine learning algorithm employed for the prediction of continuous values. This algorithm models the relationship between explanatory variables and the target variable through a straight line that best describes their association. Multivariable linear regression

(MLR) simultaneously considers two or more explanatory variables for a single target variable, thereby enabling the incorporation of a greater number of factors compared to simple linear regression. Through this approach, path loss modeling can be provided that considers the relative influence of each propagation environment variable.

B. Support Vector Regression

Support Vector Regression (SVR) solves regression problems by identifying a function that best fits the data while keeping prediction errors within an acceptable tolerance margin. In addition, SVR employs kernel functions to reorganize the input so that complex attenuation patterns arising from variations in distance and frequency become more separable, allowing the model to represent nonlinear changes in path loss. This capability is particularly beneficial for modeling real-world propagation data, where irregular behaviors caused by multipath fading, diffraction, scattering, and shadowing frequently appear. As a result, SVR provides a flexible and practical approach for capturing propagation characteristics that cannot be sufficiently described by simple linear models.

C. Multi-Layer Perceptron

Artificial Neural Networks (ANN) are computational models inspired by the way biological neurons process information, enabling them to learn nonlinear relationships within data for tasks such as prediction and classification. In this work, a Multi-Layer Perceptron (MLP) was adopted as the ANN-based model. The MLP is a typical feedforward architecture composed of several hidden layers positioned between the input and output layers. Each neuron passes information through weighted connections and nonlinear activation functions, allowing the network to represent and learn complex patterns embedded in the dataset.

III. Training Process

A. Data Preprocessing

The dataset used in this study consists of three variables: frequency (MHz) at 3400, 5300, and 6400 MHz, distance (m), and path loss (dB). According to radio propagation theory as confirmed in the free-space path loss (FSPL) model path loss can be expressed linearly when frequency and distance are represented on a log scale, as in [4]:

$$\mathrm{FSPL_{dB}} = 20\log_{10}\left(\frac{4\pi df}{c}\right) \tag{1}$$

Therefore, in this study, log scale variables for frequency and distance were generated and the primary explanatory variables are d, f, where d represents the 2D-distance in meters, and f represents the carrier frequency in MHz. Furthermore, as shown in Eq. (1), path loss (dB) is expressed as log scale terms of frequency and distance, suggesting that log-distance and log-frequency are more suitable expressions for describing path loss. Therefore, this paper ultimately adopts $log(d)$ and $log(f)$ as explanatory variables for modeling.

TABLE I
THE HYPERPARAMETERS VALUE USED FOR EACH MODEL

Model	**Hyper Parameter**	**Area A**	**Area B**
MLR	-	-	-
SVR	kernel	rbf	rbf
	C	1	3
	gamma	0.05	1.0
	epsilon	0.1	0.5
MLP	hidden layer & node	[25, 25, 25]	[25, 25, 25]
	learning rate	0.01	0.01
	activation function	Layer1: tanh Layer2: sigmoid Layer3: sigmoid	

B. Learning Process

In this study, path loss was predicted utilizing MLR, SVR, and MLP models. To this end, the entire dataset was partitioned into training (70%) and testing (30%) subsets to evaluate the generalization performance of the models. Furthermore, all variables were normalized to a 0-1 range to reduce scale differences between variables and prevent any specific variable from exerting excessive influence during the learning process.

TABLE I represents the optimal hyperparameters selected during the predictive modeling process. In the hyperparameter search process, 20% of the total test data was used as validation data, and the combination with the highest average R^2 performance through 5-fold cross-validation was determined as the final parameters. MLR has no hyperparameters since the regression coefficients are directly determined through training. Meanwhile, for MLP, the number of nodes was varied within the range of 3 to 65 to compare performance, and since performance improvement was not significant beyond 20-25 nodes, it was fixed at 25. For activation functions, a total of 27 cases combining ReLU, tanh, and sigmoid for each layer were evaluated, and the combination showing the best R^2 value was finally adopted.

IV. Simulation Result

A. Regression Result

Based on empirically measured path loss data, path loss prediction as a function of distance and frequency was conducted using MLR, SVR, and MLP models. The performance of each model was evaluated using test data with RMSE and R^2 employed as the primary evaluation metrics. TABLE II presents a summary of the prediction performance for each model.

TABLE II
Performance Metrics for Each Model

	Area A		Area B	
Method	R^2	RMSE [dB]	R^2	RMSE [dB]
MLR	0.5205	7.290	0.7538	7.139
SVR	0.5169	7.405	0.8124	6.232
MLP	0.5368	7.118	0.8221	6.041

A comparative analysis of the predictive performance of MLR, SVR, and MLP on measured path loss data in Area A and Area B revealed that MLP demonstrated the most superior performance overall, exhibiting the highest R^2 values and the lowest RMSE.

In Area A, the differences in R^2 and RMSE among MLR, SVR, and MLP were not substantial, with all three models showing comparable predictive performance at approximately $R^2 \approx 0.52 \sim 0.53$ and RMSE levels of 7.1–7.4 dB. This can be interpreted as indicating that the path loss characteristics in Region A are relatively simple with strong linearity, thereby limiting the pronounced advantages of nonlinear models. However, as demonstrated in Fig. 1 and Fig. 2, certain distance intervals exhibit a "two-stage increase pattern" where path loss increases rapidly, then temporarily levels off before increasing again, which cannot be adequately approximated by linear models. In contrast, the MLP follows these complex variation patterns with greater precision and demonstrates the most superior results in performance metrics, clearly manifesting its effectiveness.

Conversely, Area B exhibited distinct performance differences among the models. While MLR achieved $R^2 \approx 0.75$ with an RMSE of approximately 7.1 dB, SVR and MLP attained $R^2 \approx 0.81, 0.82$ and RMSE values of approximately 6.2 dB and 6.0 dB, respectively, thereby significantly reducing errors compared to linear regression. This can be attributed to the fact that path loss data in the region B exhibits strong nonlinearity and high variability due to building distribution and scattering effects, enabling nonlinear models to more effectively learn these complex patterns. Furthermore, as can be observed in Fig 1 and 2, the differences between the predictive curves of the models are clearly evident in Area B as well. While MLR follows the overall increasing trend, it fails to adequately capture the fluctuations and variability patterns present in the observed data. In contrast, SVR and MLP track these variations more precisely, with MLP in particular demonstrating a tendency to reproduce the flow of data distribution most naturally, thereby exhibiting the most superior fit visually.

In summary, while linear regression serves as a meaningful baseline model, nonlinear models such as SVR and MLP significantly enhance path loss prediction accuracy as environmental complexity increases, and as evidenced in Fig. 1 and Fig. 2, nonlinear models demonstrate substantially more precise reflection of the variation trends observed in measured data compared to linear regression.

V. Conclusion

In this study, the predictive performance of MLR, SVR, and MLP models was comparatively analyzed utilizing field measurement-based propagation path loss data. The analysis revealed that all three models exhibited similar performance in regions Area A; however, in regions characterized by strong nonlinear properties, SVR and MLP recorded higher R^2 values and lower RMSE compared to MLR, demonstrating superior predictive performance. In particular, the MLP not only achieved the highest accuracy but also effectively captured subtle nonlinear variations present in the data even in regions of strong linearity, clearly demonstrating the necessity of nonlinear models in complex propagation environments. This study demonstrated that AI modeling approaches based on empirical measurement data are effective for propagation prediction, and future research may extend toward more sophisticated propagation prediction models through data diversification, incorporation of spatial and environmental information such as satellite imagery-based building distribution and advancement of deep learning architectures.

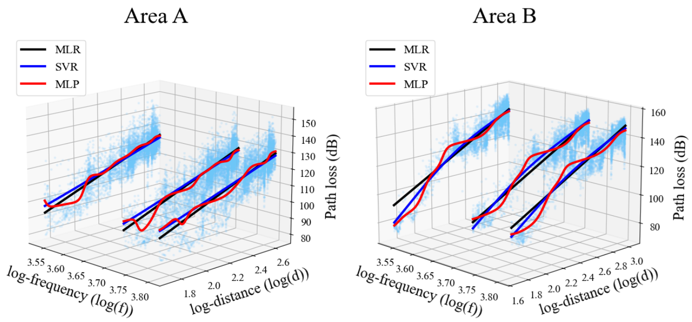

Fig. 1. 3D Path Loss Prediction Results at 3400, 5300, and 6400 MHz for Area A and Area B

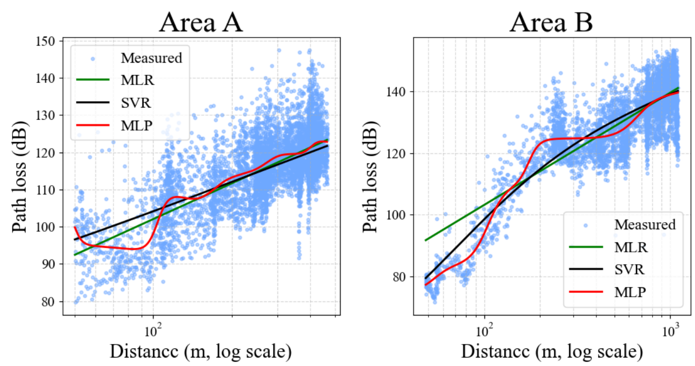

Fig. 2. Path Loss Prediction Results at 3400 MHz for Area A and Area B

Acknowledgment

This work was supported by the National Research Foundation of Korea (NRF) Grants funded by Korean Government through Ministry of Science and ICT (MSIT) under Grant RS-2024-00409492 and Grant RS-2024-00346319.

References

[1] ITU-R Question 236/3, "Use of machine learning methods for radio-wave propagation studies," 2023.

[2] H. S. Jo and J. G. Yook, "Path loss characteristics for IMT-advanced systems in residential and street environments," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 867–871, 2010.

[3] COST Action 231, "Digital mobile radio towards future generation systems, final report," *Tech. Rep.*, European Communities, EUR 18957, 1999.

[4] ITU-R P.525-4, "Calculation of free-space attenuation," *ITU*, Geneva, 2019.