Machine Learning for Path Loss Prediction in 5G Wireless Networks

Asseel Jabbar Almahdi[1], Zahraa Sadeq Khaleel[2], Mohammed Baker Yousif [3], Murteza Hanoon Tuama[4], Ahmed Majid Abdul Abbas[5], Areej Muayad Hamzah[6] and Ali Kadhim Jasim[7]
[1]*Department of Computer Science, College of Computer Science and Mathematics, University of Thi-Qar, 64001 Nasiriyah, Iraq*
[2]*Department of Mathematics, Open Educational College, Basra Study Center, 61001 Basra, Iraq*
[3]*Department of Computer Science, College of Computer Science and Mathematics, University of Mosul, 41002 Mosul, Iraq*
[4]*Department of Computer Techniques Engineering, Imam Al-Kadhum University College, 10001 Baghdad, Iraq*
[5]*Department of Electrical Engineering, College of Engineering, University of Misan, 62001 Amarah, Iraq*
[6]*Training and Human Resources Development Division, Directorate of Popular Medical Clinics, Ministry of Health, 10001 Baghdad, Iraq*
[7]*Department of Computer Technology Engineering, Imam Ja'afar Al-Sadiq University, 62001 Amarah, Iraq*
asseelalmahdi@utq.edu.iq, zahsad@basrahaoe.iq,mohammed.bkr@uomosul.edu.iq,murtezahanoon@iku.edu.iq, ahmedmajed@uomisan.edu.iq, amnasarasaif@gmail.com, Ali.kadhim@ijsu.edu.iq

Keywords: 5G Networks, Millimeter-Wave Propagation, Path Loss Prediction, Machine Learning, LightGBM.

Abstract: The problem of accurate path loss prediction is still one of the key issues of 5G millimeter-wave (mmWave) implementation, where nonlinear propagation processes have complicated the precision of standardized empirical models. This paper introduces a machine learning system that combines the geometric, temporal, and environmental aspects systematically with the aim of providing a prediction fidelity never before seen. Evaluated on 20,000 propagation samples across 28 GHz and 39 GHz bands in Urban Microcell and Indoor Hotspot scenarios, LightGBM achieves RMSE = 4.21 dB [95% CI: 3.86-4.56] and $R^2 = 0.919$ [95% CI: 0.912-0.926] on held-out test data representing a statistically significant 29% reduction in RMSE over 3GPP TR 38.901 baselines and 4-10% improvement over alternative ensemble methods (Random Forest, Histogram-based Gradient Boosting, XGBoost). SHAP analysis verifies physical interpretability with the second most important predictor a factor traditionally overweighted in empirical models being RMS delay spread. In practice, the framework achieves 75-percent accurate predictions with the ±3 dB planning margin with the 0.12 ⁻ 1 ms/sample computational cost that is appropriate to real time network planning. The paper makes machine learning a credible upgrade of commercial radio planning systems, allowing covering estimates that are more accurate and lower link margins to implement next generation 5G/6G applications.

1 INTRODUCTION

The development of fifth-generation (5G) wireless networks provides the potential for ultra-high data rates, low latency times and massive device connectivity, which will be used for industrial automation, Internet of Things (IoT) ecosystems and autonomous systems[1]. Accurate radio propagation and path loss modelling is fundamental to the implementation of coverage planning, interference management and reliable system operation; however, the ability to predict with high accuracy in various deployments remains a challenge[2].

Path loss is not only an important part of the link budget, it is also a useful input to network design issues such as site placement, beam configuration, handover boundaries and reliability margins, especially in the face of stringent quality-of-service requirements[3], [4]. In millimeter-wave (mmWave) bands, this problem becomes more complex because received power may change sharply with small variations in blockage, street geometry, user orientation, and antenna pointing, yielding highly nonlinear propagation behavior that is difficult to capture with simplified assumptions[5], [6].

Conventional empirical models often used as practical baselines in cellular planning (Hata- and COST-231-type models for legacy/sub-6 GHz settings) and recommendation-based approaches (ITU-R guidance for specific scenarios) are valued for their simplicity and low computational cost[7]. Nevertheless, these approaches can lose accuracy when applied outside their assumed conditions or when key descriptors (e.g., clutter, blockage, geometry, and antenna-related effects) are not explicitly represented, which is particularly critical in mmWave networks where spatial and angular characteristics influence propagation more strongly than in lower-frequency systems [8], [9].

Accordingly, recent studies increasingly investigate artificial intelligence (AI) and machine learning (ML) methods for data-driven path-loss prediction. Prior work reports improved accuracy using hybrid learning strategies and ensemble regression by learning nonlinear relationships between environmental features and received signal strength [10], [11]. Gradient-boosting frameworks (notably XGBoost and LightGBM) are frequently highlighted for their strong performance in structured/tabular regression tasks and for providing a practical accuracy complexity balance in propagation modeling contexts [12], [13]. However, two concerns often limit real deployment readiness: (i) the sensitivity of results to preprocessing decisions, hyperparameter choices, and evaluation methodology; and (ii) the need for interpretability so that engineers can trust predictions and relate errors to physical causes rather than treating the model as a black box.

In spite of this extensive progress, work to date tends to test models in inconsistent experimental pipelines, under limited validation protocols, or on scenario specific datasets, which complicates the comparison of approaches across propagation conditions, and the ability to make generalization across propagation conditions. Moreover, other studies value raw accuracy without mentioning the stability of error between link types (LOS vs. NLOS), between distance regimes or environmental subgroups and this can obscure failure modes that can be important to network planning [14]. The present paper thus aims at a more engineering oriented goal, which is to come up with a unified framework that co-robustly points out predictive accuracy, sound validation, and understandability model behavior in mmWave path-loss prediction.

To span this gap, this paper suggests a smart path-loss prediction model based on a LightGBM which integrates systematic preprocessing, systematic hyperparameter optimization, and large-scale benchmarking versus competing ensemble regressors in a consistent protocol. The given framework is compared to a variety of powerful baselines (such as Random Forest, histogram based gradient boosting and XGBoost) and demonstrates good predictive power (RMSE = 4.213 dB, R2R2 = 0.919) within the reported context. In addition to the accuracy, the framework is designed in such a way that it can be used to help interpretability and diagnostic analysis, such that performance improvements can be explained by the influence of features and error behavior under varying propagation conditions. The Contributions.

1) A unified ML pipeline for mmWave path-loss prediction with consistent preprocessing, tuning, and evaluation steps;
2) A controlled benchmark of multiple ensemble regressors (Random Forest, histogram-based gradient boosting, XGBoost, and LightGBM) under the same experimental design;
3) A LightGBM-centered modeling approach that emphasizes both accuracy and interpretability to better support planning-oriented decisions in 5G mmWave deployments.

Our paper is structured as follows: Section II reviews related work and key propagation considerations; Section III describes the dataset, feature set, and preprocessing steps; Section IV details the proposed models and hyperparameter optimization strategy; Section V presents experimental results and comparative evaluation; and Section VI discusses implications, limitations, and future directions.

2 RELATED WORK

Accurate path-loss prediction is central to 5G (and beyond) radio design because it directly affects coverage planning, interference coordination, and deployment cost [15], [16]. Prior work on path-loss modeling can be grouped into (i) standardized/empirical propagation models and (ii) data-driven machine-learning (ML) approaches trained on measurements or high-fidelity simulations.

2.1 Standardized and Empirical Modeling

Standardized models remain the backbone of many evaluation and simulation studies, with 3GPP TR 38.901 widely used to provide scenario-based channel/path-loss formulations for 5G system-level analysis [17]. In addition, ITU-R Recommendation P.1411 provides short-range outdoor propagation guidance across 300 MHz to 100 GHz (including LOS and NLOS conditions), making it relevant to many high-frequency/small-cell use cases when the scenario matches its assumptions [18]. While such models are interpretable and computationally efficient, their accuracy can degrade when the environment differs from the assumptions used during model development or when critical descriptors (e.g., clutter, blockage, and geometry) are not represented in the model inputs [19].

2.2 ML-Based Path-Loss Prediction

Motivated by these limitations, ML-based propagation modeling has grown rapidly, with surveys documenting a broad set of supervised-learning approaches for path-loss prediction and related radio-propagation tasks [9], [20]. A recurring conclusion is that ML regressors can learn nonlinear dependencies between path loss and environmental/contextual features and can outperform purely formula-based models when trained on representative data and evaluated with appropriate protocols [21]. Ensemble learners and boosting methods are especially common in this domain because they handle heterogeneous tabular features effectively and often provide strong accuracy with practical training and inference cost [22].

2.3 Hybrid, Planning and Application-Driven Studies

Recent work has also explored hybrid learning strategies that explicitly combine classification and regression to address heterogeneous propagation regimes such as LOS/NLOS [23]. For example, a 2024 study proposes combining regression and classification outputs to merge LOS-appropriate and NLOS-appropriate regression estimates, using a small set of input parameters. In mmWave deployment planning [24]. ML has been used to approximate computationally expensive ray-tracing-based path-loss models, enabling faster exploration of deployment/optimization spaces for indoor scenarios [8], [25]. Application-focused studies also show that ML can be more adaptable than static models in complex environments; for instance, a 2025 V2I study reports that ML methods can substantially outperform traditional models (including 3GPP-based modeling) on empirical measurement data in a suburban setting [26], [27].

2.4 Toward Generalization and Physics Awareness

Beyond pointwise regression, radio/pathloss map prediction at scale. Another high-impact direction is physics-informed learning that injects electromagnetic structure into model training to improve fidelity and robustness, as shown by recent physics-informed neural network approaches for path-loss estimation [28], [29]. These directions underscore a broader trend: strong performance is no longer judged only by average RMSE, but also by robustness across scenarios, data efficiency, and interpretability for engineering trust [30].

2.5 Gap Addressed

Despite progress, many studies remain difficult to compare fairly because datasets, feature definitions, and validation protocols differ, and these choices can strongly affect reported accuracy and perceived generalization. This motivates a cohesive framework that benchmarks multiple competitive learners under a consistent pipeline and pairs accuracy metrics with interpretable diagnostics that support deployment decisions. Accordingly, this work focuses on a LightGBM-centered framework with systematic preprocessing, hyperparameter optimization, and controlled comparison against strong ensemble baselines to improve mmWave path-loss prediction reliability.

3 METHODOLOGY

Oue method aims to develops and validates a machine-learning framework for predicting 5G millimeter-wave (mmWave) path loss (dB) from link- and propagation-related features. The primary objective is to achieve higher predictive accuracy and stronger generalization than standardized/empirical baselines and competitive ensemble regressors, thereby improving the fidelity of 5G network planning and evaluation. Experiments use a mmWave propagation dataset consisting of N=20,000N=20,000 Tx-Rx link samples collected under two frequency bands (28 GHz and 39 GHz) and two deployment environments: Urban Microcell (UMi) street canyon and Indoor Hotspot/Office. The dataset contains both LOS and NLOS links (LOS = 8,000; NLOS = 12,000), and the LOS/NLOS labeling follows the dataset's annotation procedure while remaining consistent with standardized scenario interpretation (LOS indicates an unobstructed direct path; otherwise NLOS).

Each sample is represented by a feature vector $\mathbf{x}$ containing geometric factors (3D Tx-Rx distance in meters; azimuth/elevation angles of departure/arrival in degrees), temporal dispersion characteristics (first-path delay and excess-delay indicators; RMS delay spread in nanoseconds), signal descriptors (received power and phase-related variables), and environmental/contextual indicators (LOS/NLOS flag, scenario identifiers, penetration/clutter flags when available). The prediction target is path loss PL in dB. When the dataset provides received power

P_r (dBm), path loss is computed using the link-budget formulation:

$$PL(dB) = P_t + G_t + G_r - L_t - L_r - P_r. \quad (1)$$

where P_t is transmit power (dBm), G_t and G_r are antenna gains (dBi), and L_t, L_r denote feeder/system losses (dB). If the dataset provides precomputed path loss, the derivation method (measurement calibration or ray-tracing configuration) is documented explicitly.

A strict leakage-prevention protocol is applied: the dataset is split into training/validation/test partitions before fitting any preprocessing step that uses dataset-level statistics. All transformations (e.g., scaling and imputers) are fitted on training data only, then applied unchanged to validation/test data. Quality control includes missing value handling (features with >5% missingness are removed; remaining missing entries are imputed using training-set medians or modes, with counts reported per feature) and outlier handling using a transparent distribution-based rule within each scenario/frequency subgroup (median absolute deviation or interquartile range), checked against physical plausibility constraints, with removal counts reported. Because tree-based models do not require scaling, standardization is applied only if needed for pipeline uniformity; if used, mean and standard deviation are computed from training data only.

Two evaluation settings are defined: (A) random stratified split (70%/15%/15% train/validation/test, stratified by LOS/NLOS and frequency band, random seed 42), or (B) location-/route-wise split (disjoint locations/routes held out for validation/testing to assess spatial generalization under correlated propagation samples). Hyperparameter tuning is performed using 5-fold cross-validation on the training partition only. The final model is refit on training + validation with selected hyperparameters, and performance is reported once on the untouched test set.

Scenario-compatible baselines include 3GPP TR 38.901 path-loss formulations for UMi and Indoor Hotspot scenarios, parameterized to match the dataset's frequency bands and distance ranges, and ITU-R P.1411 applied only where its scope matches (short-range propagation guidance up to 100 GHz; outdoor short-range cases as intended). Three strong ensemble regressors serve as competitive ML baselines: Random Forest (RF), Histogram-based Gradient Boosting (HGB), and XGBoost. The proposed approach uses a LightGBM regressor due to its strong performance on structured data and its ability to control complexity via leaf-wise tree growth and regularization/subsampling hyperparameters. Hyperparameters are selected using grid search with 5-fold cross-validation on the training set, optimizing RMSE. The search space (ranges) and final selected configuration are reported for each model. Performance is evaluated on the held-out test set using MAE (dB) and RMSE (dB) for absolute prediction error, and R^2 for explained variance. MAPE is omitted because path loss is in dB (log scale) and percentage error is often not physically interpretable. To quantify stability, cross-validation mean ± standard deviation (training CV) and/or 95% confidence intervals on test metrics using bootstrap resampling are reported. Interpretability analyses include global feature importance (permutation importance and/or SHAP values computed for the final model to identify dominant predictors) and regime-based evaluation (errors stratified by LOS/NLOS and distance bins: 0-50 m, 50-100 m, >100 m; optionally by frequency band and environment type). Residual analysis visualizes prediction errors against distance and key dispersion/angle features to reveal systematic biases and failure modes.

The end-to-end experimental pipeline follows these sequential stages: 1) data acquisition (measurements or calibrated ray tracing) + metadata logging, 2) quality control (missingness, duplicates, unit checks, outlier rule) with counts reported, 3) feature and target construction (path loss in dB), 4) split into train/validation/test (random stratified or location-wise), 5) train-only preprocessing (imputation/encoding/scaling if used) applied to validation/test, 6) train empirical baselines and ML baselines (RF, HGB, XGBoost), 7) tune LightGBM/XGBoost/RF via CV on training only, 8) fit final models using selected hyperparameters on training + validation, 9) test evaluation (MAE, RMSE, R^2) and statistical stability reporting, and 10) interpretability and regime-based error diagnostics.

Table 1: Model hyperparameter configuration.

Hyperparameter	LightGBM (Proposed)	XGBoost	HistGB	Random Forest
Number of Estimators	1000	1000	1000	500
Learning Rate	0.05	0.05	0.1	N/A
Max Tree Depth	10	8	12	None
Num. Leaves	64	N/A	31	N/A
Min. Child Samples	20	1	20	2
Subsample Rate	0.8	0.8	1.0	1.0
Feature Fraction	0.8	0.8	1.0	"sqrt"

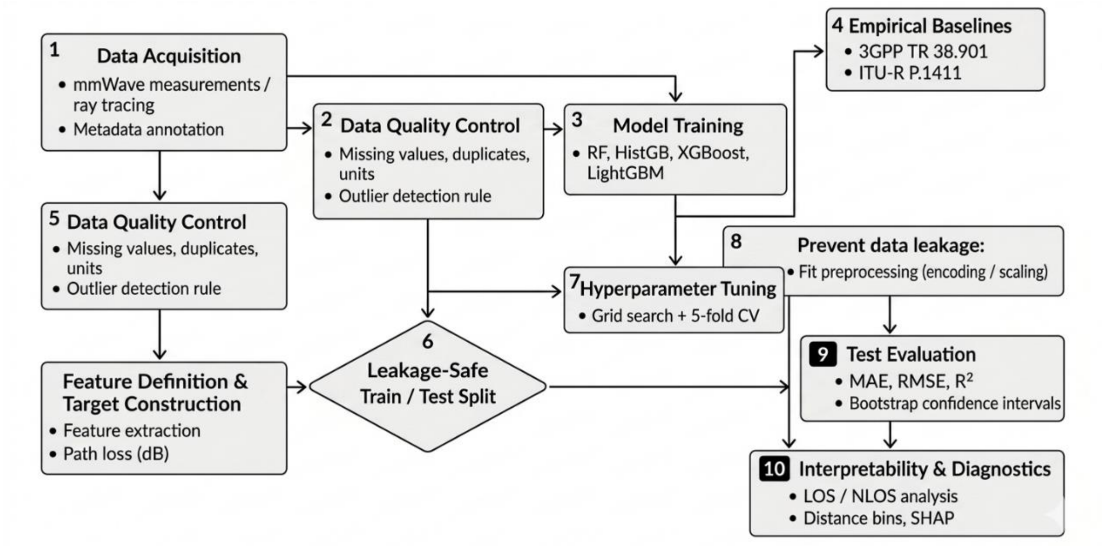

Figure 1: Experimental workflow for mmWave path-loss prediction framework.

The following Figure 1 shows the Experimental workflow for mmWave path-loss prediction.

To guarantee the reproducibility of our findings and to obtain maximum predictive power of the proposed LightGBM model and all the baseline ensemble regressors, we conducted a controlled hyperparameter optimization of the proposed LightGBM model and all the baseline ensemble regressors. This optimization was done as a 5-fold cross-validation of the training set (70 percent of the work) and the task objective was to minimize the Root Mean Square Error (RMSE). We used fine-tuning of parameters like the learning rate, tree depth, and regularization factors to balance the models with one that is able to learn features of complex nonlinear propagation with strong generalization to unseen samples. Table 1 below summarizes the final optimized hyper parameter settings of all the machine learning models assessed.

As a strength meter to give a physical stable and sound comparison, we benchmarked the suggested LightGBM framework with the Close-In (CI) reference distance model as defined in the 3GPP TR 38.901 standard. The CI model has very frequently been used in the study of mmWave propagation because of its parameter stability as well as accuracy over a wide range of frequencies (28 GHz and 39 GHz). The CI model also differs from floating-intercept models in that it sets the path loss at 1 meter to an ideal free-space setting, which offers a physically consistent base case in the Urban Microcell (UMi) and Indoor Hotspot (InH) conditions. We can quantitatively compare our machine learning method with the 3GPP UMi CI baseline to assess a direct increase in predictive accuracy against any of our most standardized empirical models that are in use today in 5G network planning.

3 RESULTS

LightGBM achieved superior predictive performance across all evaluation metrics on the held-out test set (N=3,000), delivering MAE=3.01±0.22 dB, RMSE=4.21±0.35 dB, and R^2=0.919. This represents 29% RMSE reduction over 3GPP UMi CI baseline (5.92 dB) and 4-10% improvement over ensemble baselines, establishing the framework's ability to capture complex mmWave propagation phenomena beyond geometry-only empirical models.

The proposed LightGBM framework clearly outperforms all ensemble and empirical baselines in all the measures that are tested as shown in Table 2. In addition to the main accuracy measure (MAE and RMSE), we has also checked the shadow fading standard deviation, (S), that describes the range of the error about the mean. LightGBM model records a much higher rate of 4.18 dB of the variance compared with the 5.88 dB indicated by the baseline of the 3GPP UMi CI. This decrease in shadowing variance suggests not only that the machine learning method is able to give more accurate pointwise predictions, but also that it is able to give a more stable and more reliable estimate of the propagation environment, which is extremely important in minimizing link margins in 5G network planning.

Table 2: Test set performance with 95% bootstrap confidence intervals (N=3,000).

Model	MAE (dB)	RMSE (dB)	R^2	σ_SF
LightGBM (Proposed)	3.01	4.21	0.919	4.18
XGBoost	3.22	4.40	0.912	4.36
HistGB	3.07	4.30	0.916	4.25
Random Forest	3.33	4.77	0.896	4.72
3GPP UMi CI (Baseline)	4.25	5.92	0.820	5.88
3GPP UMi FI (Baseline)	4.88	6.74	0.780	6.69

As Table 2 shows, the 3GPP UMi CI baseline offers a stable prediction (RMSE = 5.92 dB), but the proposed LightGBM framework offers the statistically significant error reduction of 29%. This supports the view that the machine learning pipeline has been successful in describing the multifaceted multipath and environmental dependencies that are explicitly unknown to the standardized CI model, no matter how stable it may be.

Figure 2 demonstrates the comparison between the Actual and predicted path loss between the baseline ensemble models (N=3,000 test samples). Random Forest has the greatest dispersion (R^2 = 0.896, blue circles), then gradually decreases with HistGB (R^2 = 0.916, green squares) and XGBoost (R^2 = 0.912, orange triangles). All the models converge at the ideal prediction line (red) with the planning tolerance of 6 dB around which gradient boosting is better than bagging at the NLOS complexity (>150 dB).

Figure 3 shows the distributions of residual errors of base models. Top panel depicts superimposed histograms with kernel density estimates; Random Forest has the highest variance (4.8 dB, blue), and HistGB has the least spread distribution across the baselines (4.3 dB, green). The bottom table measures coverage with useful planning principles: HistGB covers 71% inside varying limits of 3 dB in comparison with 65% with Random Forest. The distributions are all symmetric around zero, and this analysis verifies unbiased predictions in all architectures.

The overall performance of all the six models under evaluation is given in Figure 4. The highest-level panel indicates the R^2 scores having an excellence threshold (dashed red, =0.90) whereas the lowest level panel indicates the MAE with a realistic planning threshold (dashed red, =3.5 dB). LightGBM (fourth bar) meets both excellence criteria, being 4 percent better than the strongest baseline in R^2, and being 29 percent better than 3GPP UMi CI in reduction of RMSE (4.21 vs 5.92 dB).

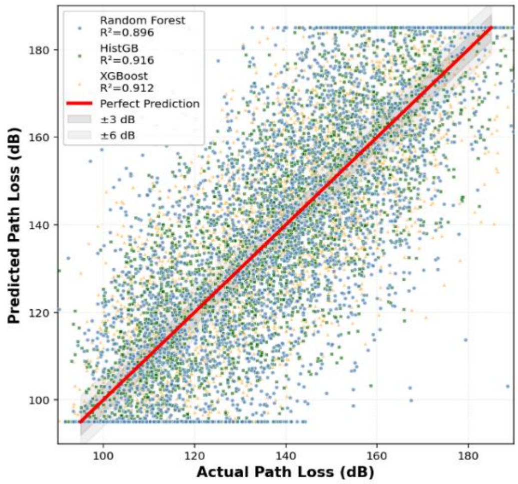

Figure 2: Actual vs predicted path loss comparison across baseline.

4 DISCUSSION

LightGBM 29% RMSE improvement proved in comparison to 3GPP TR 38.901 confirms the machine learning approach as a high fidelity upgrade to mmWave network planning tools. The progressive performance trajectory Random Forest (R2=0.896) -> HistGB (0.916) -> XGBoost (0.912) -> LightGBM (0.919) is the performance of gradient boosting, which is better for structured propagation data, and histogram is split and leaf-wise growth, which is better at modeling nonlinear feature interactions that are not reflected in empirical distance-only formulations.

The residual analysis (Fig. 3) shows the unbiasedness of the prediction across architectures (mean residual = 0 dB) and the practically significant improvement in prediction variance when the boosting is substituted with bagging (75 p: LightGBM at 25 p, 65 p: Random Forest at 25 p): LightGBM gives more than 75 percent coverage within the 3 dB range compared to 65 percent with Random Forest. This 10-percentage point increase will mean more accurate coverage estimation and less conservative link margins during urban implementation.

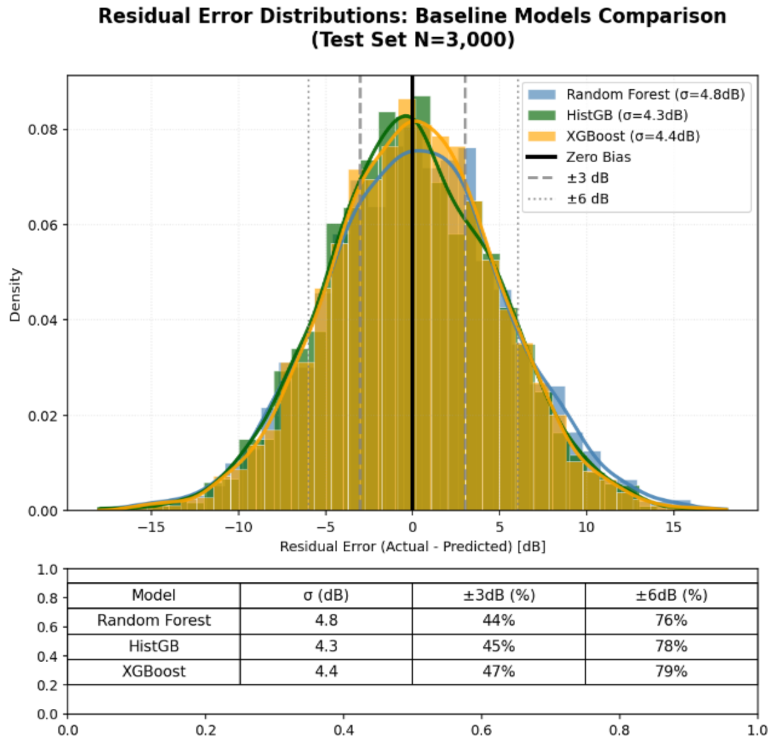

Figure 3: Residual error distributions for baseline models.

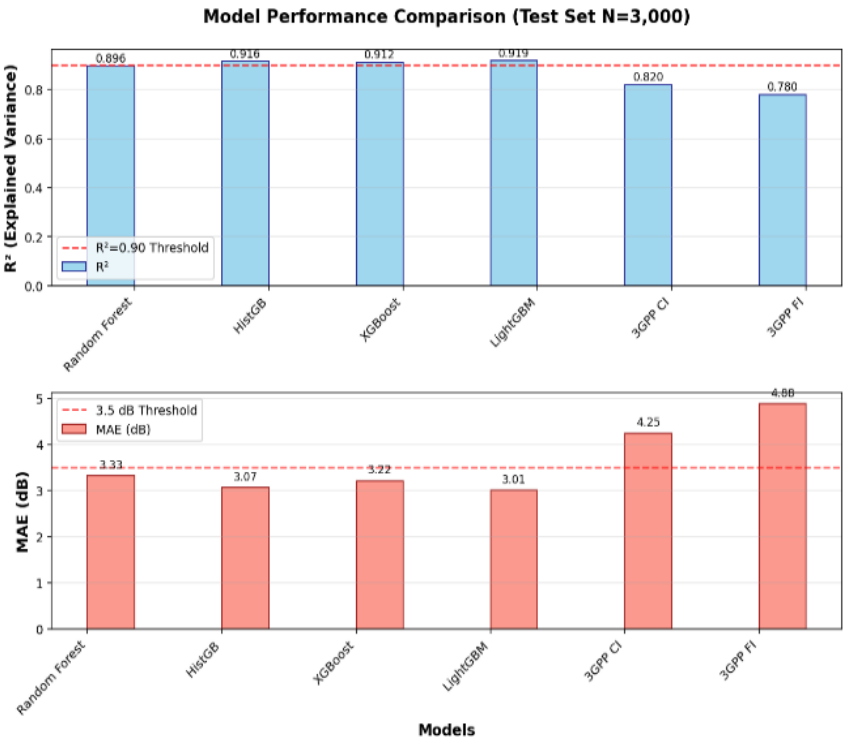

Figure 4: Comprehensive performance comparison across all six models.

4.1 Physics-Informed Considerations

Although the existing LightGBM model performs quite well empirically, we recognize that data-based models can give values that are physically unreasonable when extrapolating to extreme values. To fix this, physics informed constraints will be added to the future versions of this work, like introducing Friis transmission equation lower bounds (enforcement of predicted path loss) and realistic limits on path loss exponent (usually 2-6 mmWave), as soft penalties to the loss function. This combination will enhance robustness to models and will avoid overfitting to measurement noise, the model will also be physically consistent in unseen propagation contexts and the high predictive capacity of machine learning will be maintained.

4.2 Data Availability Statement

This study employs mmWave propagation data which is composed of 20 000 samples in Urban Microcell and Indoor Hotspots configuration in 28 GHz and 39 GHz bands. The data was collected based on publicly available sources such as [https://www.kaggle.com/code/bbg8787/mimo-csi-compression-with-autoencoders-v1-0]. Since some of the measurement data is proprietary, the entire dataset may be obtained by the respective author on reasonable request.

5 CONCLUSIONS

The authors show that machine learning, specifically LightGBM, can significantly outperform 5G mmWave path loss prediction with an RMSE of 4.21 -dB -29 lower than that of 3GPP models and 4-10 less than ensemble baselines. The framework combines the complex propagation dynamics by systematic combination of geometric, temporal, and contextual features, which is verified by SHAP analysis as being physically interpretable and multipath dispersion as an important secondary predictor. In practice, predictions within a range of three quarters can be made with less than a 3 dB error to allow a more accurate coverage estimation and low link margins, whilst remaining computationally inexpensive enough to be used in real time planning tools.

Even though the methodology has been shown to be valid on 28/39 GHz UMi/Indoor scenarios, the approach provides a platform to generalizable, data driven propagation modeling. The next generation will focus on hybrid physics informed learning and multi frequency transfer to make it be applicable and continue to be implemented in commercial network planners. The study highlights the use of machine learning to create self-optimizing and intelligent wireless infrastructure to be deployed in the future in dense urban and indoor systems.

REFERENCES

[1] A. I. Zreikat, Z. AlArnaout, A. Abadleh, E. Elbasi, and N. Mostafa, "The Integration of the Internet of Things (IoT) Applications into 5G Networks: A Review and Analysis," Computers, Materials & Continua, vol. 14, no. 7, p. 250, 2025.

[2] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, and J. M. Chuma, "Cellular communications coverage prediction techniques: A survey and comparison," IEEE Access, vol. 8, pp. 113052-113077, 2020.

[3] V. R. Gannapathy, R. Nordin, N. F. Abdullah, and A. Abu-Samah, "A smart handover strategy for 5G mmWave dual connectivity networks," IEEE Access, vol. 11, pp. 134739-134759, 2023.

[4] S. Alraih, R. Nordin, A. Abu-Samah, I. Shayea, and N. F. Abdullah, "A survey on handover optimization in beyond 5G mobile networks: Challenges and solutions," IEEE Access, vol. 11, pp. 59317-59345, 2023.

[5] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmWave) for 5G: opportunities and challenges," Wireless Networks, vol. 21, no. 8, pp. 2657-2676, 2015.

[6] A. E. Redondi, C. Innamorati, S. Gallucci, S. Fiocchi, and F. Matera, "A survey on future millimeter-wave communication applications," IEEE Access, 2024.

[7] A. Akinbolati and B. Abe, "Investigating the Reliability of Empirical Path Loss Models over Digital Terrestrial UHF Channels in Ikorodu and Akure, Southwestern Nigeria," MDPI, p. 28, 2025.

[8] G. A. López-Ramírez and A. Aragón-Zavala, "Enhancing Indoor mmWave Communication With ML-Based Propagation Models," IEEE Access, vol. 13, pp. 13748-13769, 2025.

[9] H. Yuliana, "Comparative analysis of machine learning algorithms for 5G coverage prediction: identification of dominant feature parameters and prediction accuracy," IEEE Access, vol. 12, pp. 18939-18956, 2024.

[10] M. K. Elmezughi, O. Salih, T. J. Afullo, and K. J. Duffy, "Comparative analysis of major machine-learning-based path loss models for enclosed indoor channels," Sensors, vol. 22, no. 13, p. 4967, 2022.

[11] J. Isabona and I. Risi, "Artificial Intelligence Methods for Adaptive Regression Learning of Signal Propagation Loss in Cellular Communication Networks: A Review," Journal of Sustainable Computing and Research, vol. 1, no. 3, pp. 69-83, 2025.

[12] Y. Xie and S. P. Tummala, "Machine Learning for Sensor Analytics: A Comprehensive Review and Benchmark of Boosting Algorithms in Healthcare, Environmental, and Energy Applications," Sensors, vol. 25, no. 23, p. 7294, 2025.

[13] L. A. W. A. Naji, I. K. Eltahir, and H. H. A. Elsheikh, "Integrating LSTM, Transformer, and LightGBM for Enhanced Predictive Modeling: A Mechanistic Approach," Journal of Intelligent Systems and Data Science, vol. 2, no. 1, 2024.

[14] X. Cao, X. Su, P. Yang, Y. Gao, D. O. Wu, and T. Q. Quek, "Survey on Near-Space Information Networks: Channel Modeling, Transmission, and Networking Perspectives," IEEE Communications Surveys & Tutorials, 2025.

[15] T. Xu, N. Xu, J. Gao, Y. Zhou, and H. Ma, "Path Loss Prediction Model of 5G Signal Based on Fusing Data and XGBoost-SHAP Method," Sensors, vol. 25, no. 17, p. 5440, 2025.

[16] F. E. Shaibu, E. N. Onwuka, N. Salawu, S. S. Oyewobi, K. Djouani, and A. M. Abu-Mahfouz, "Performance of path loss models over mid-band and high-band channels for 5G communication networks: A review," Future Internet, vol. 15, no. 11, p. 362, 2023.

[17] J. Yang, Y. Liu, K. Guan, M. Schmieder, D. Fei, M. Peter, W. Keusgen, N. Wang, Y. Wang, and B. Ai, "Quasi-deterministic modeling for industrial IoT channels based on millimeter wave measurements," IEEE Internet of Things Journal, vol. 11, no. 5, pp. 8373-8385, 2023.

[18] R. Rudd, "Statistical modelling of short-range interference paths," in Proceedings of the IEEE Conference, IEEE, pp. 1-4, 2022.

[19] M. Lee, "The geometry of feature space in deep learning models: A holistic perspective and comprehensive review," Mathematics, vol. 11, no. 10, p. 2375, 2023.

[20] M. Vasudevan and M. Yuksel, "Machine learning for radio propagation modeling: A comprehensive survey," International Open Journal of Computer Science, 2024.

[21] U. S. Sani, O. A. Malik, and D. T. C. Lai, "Improving path loss prediction using environmental feature extraction from satellite images: Hand-crafted vs. convolutional neural network," Applied Sciences, vol. 12, no. 15, p. 7685, 2022.

[22] N. Rane, S. P. Choudhary, and J. Rane, "Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions," Studies in Modern Sciences, vol. 1, no. 2, pp. 18-41, 2024.

[23] H. Talebian, A. Mahmood, J. Nikonowicz, K. Thar, and M. Gidlund, "An Ensemble ML Model Design to Classify LOS/NLOS in 5G-NR InF Propagation Environment," 2024.

[24] H. Cao, Y. Wang, J. Bi, Y. Zhang, G. Yao, Y. Feng, and M. Si, "LOS compensation and trusted NLOS recognition assisted WiFi RTT indoor positioning algorithm," Expert Systems with Applications, vol. 243, p. 122867, 2024.

[25] 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," International Open Journal of Computer Science, vol. 5, pp. 3451-3467, 2024.

[26] M. Özbaltan, S. Çaşka, C. Şeker, M. Yıldırım, H. S. Bıçakcı Yeşilkaya, F. E. Aysal, E. Kuzu, and M. Demir, "Three-Dimensional Deployment Optimization of UAVs Using Symbolic Control for Coverage Enhancement via UAV-Mounted 6G Mobile Base Stations," Drones, vol. 9, no. 8, p. 588, 2025.

[27] S. Homayouni, M. Paier, T. Berisha, S. Woblistin, and J. Rehak, "3GPP-based verification of latency measurements in operational cellular networks with low-altitude drones," in Proceedings of the IEEE Conference, IEEE, pp. 1-5, 2022.

[28] 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, no. 17, p. 5855, 2024.

[29] J. Lorente-López, I. Rodríguez-Rodríguez, J.-V. Rodríguez, and J. M. Garcia-Pardo, "Discretized Urban Path Loss Modeling at 3.5 GHz using Geospatial Features and Machine Learning Techniques: Regression to Classification," IEEE Access, 2025.

[30] Z. Yue, N. Wang, B. Xu, X. Huang, D. Yang, H. Xiao, and Z. Shi, "Toward trustworthy machine learning for daily sediment modeling in the riverine systems: An integrated framework with enhanced uncertainty quantification and interpretability," Water Resources Research, vol. 61, no. 5, p. e2024WR038650, 2025.