A Stacking Ensemble Technique to Predict Signal Path Loss via 3D GIS

Guzide Miray Perihanoglu[1], Himmet Karaman[1], Tayfun Akgül[2]

[1] ITU, Dept. of Geomatic Engineering, 80626 Maslak Istanbul, Turkey - (perihanoglu, karamanhi)@itu.edu.tr
[2] ITU, Dept. of Electronics and Communication Engineering Department, 80626 Maslak Istanbul, Turkey- tayfunakgul@itu.edu.tr

Keywords: Communication Systems, Geographic Information System (GIS), Path Loss model, Stacking Ensemble Models.

Abstract

The recent developments in cellular communication technologies, especially the emergence of 6G, have increased the need for accurate and reliable signal path loss prediction models. The accuracy of predictions is reduced because traditional empirical approaches often fail to take into account the complex relationships between radio signals and the three-dimensional urban environment. Therefore, integrating advanced machine learning algorithms with diverse geographic data offers a promising direction for improving prediction performance and supporting next-generation network planning.
This paper introduces an integrated methodology that combines Geographic Information Systems (GIS) with stacking ensemble machine learning models to enhance signal path loss prediction. The study made several key contributions, which are outlined below: (I) A GIS-based framework has been developed to integrate the Digital Twin (DT) of the study area with machine learning-based path loss models, incorporating 3D geographic data such as terrain height and building elevations. (II) The study assesses binary hybrid algorithms by examining three ensemble learning models (Gradient Boosting Machine (GBM), Extreme Gradient Boosting (XGBoost), and CatBoost). The fusion of 3D spatial data with ensemble learning algorithms has led to notable advancements in mobile network design, improving the accuracy of signal attenuation predictions. (III) Lastly, the paper emphasizes the potential of GIS-assisted machine learning techniques for future network deployments, including applications in DT, 6G, and beyond.

1. Introduction

A wide range of technologies is currently being developed to enhance the intelligence, efficiency, and sustainability of urban environments. Digital twins, which provide virtual representations of physical assets, systems, or processes updated with real-time data, (Zhang et al., 2024) are used in the monitoring, analysis, and optimisation of city components (Xia et al., 2022). This technology makes significant contributions to the sustainability and efficiency of smart cities.

Advances in communication systems are crucial for the development of smart cities and digital twin technology. High-speed internet, 5G and big data analytics are key components of these systems. However, for this infrastructure to work effectively, problems such as signal loss must be reduced (Ibhaze et al., 2016, Shaibu et al., 2023). Digital twin technology assists in predicting and optimizing communication quality by simulating propagation characteristics such as path loss and multipath effects. This simulation enables the estimation of communication quality and its subsequent optimization by utilizing signals from different frequency bands to ensure reliable communication (Cui et al., 2023). The physically accurate digital twin, scalable to city level, can be used to calculate and visualise signal quality thanks to data including the location of buildings and base stations, vegetation, elevation, radio propagation, etc (Kuruvatti et al., 2022). The extensive and detailed data sets provided by the digital twin allow machine learning algorithms to achieve higher accuracy and performance in tasks such as signal strength estimation and network optimisation.

Recently, machine learning approaches have been recognized as playing a significant part in cellular networks. Recent advances in machine learning (ML) offer a compelling alternative for applications such as mobile network functionality, location-based services and signal strength estimation. The employment of machine learning-based path loss models for the estimation of path loss has increased in prevalence when contrasted with the utilisation of empirical and deterministic path loss models. This enhancement in utilisation can be explained by the comprehensive nature of the data employed to train the machine learning models (Moraitis et al., 2021, Ojo et al., 2022). Traditional signal prediction methods are inadequate in the face of environmental complexity and dynamic variability, while machine learning provides higher accuracy and reliability thanks to its ability to learn from large data sets. Machine learning-based approaches facilitate an effective solution to overcome these challenges. Machine learning methods have the capacity to learn from large data sets and have the potential to optimise the performance of communication networks by making predictions with higher accuracy and reliability (Fauzi et al., 2022, Morocho-Cayamcela et al., 2019). In this regard, the advantages offered by machine learning-based methods have led to increasing literature interest in the area of signal path loss estimation.

Until the present day, path loss estimation studies based on ensemble methods have addressed a wide variety of stand-alone machine learning methods and ensemble methods. In some studies aimed at predicting signal path loss, artificial neural network (ANN)-based ensemble learning techniques have been compared with stand-alone machine learning methods (Kwon and Son, 2024). Utilising data obtained through image processing techniques as input, (Sotiroudis et al., 2022) designed a two-level ensemble model based on the 'stacked generalisation' approach, incorporating seven different machine learning algorithms.In other work, it has been reported that a stacking-based ensemble algorithm consisting of decision trees and random forest classifiers provides the highest accuracy for estimating the speed and distance of users (Aggarwal et al., 2024). (Goudos and Athanasiadou, 2019) has used signal strength (RSS) data collected by unmanned aerial vehicles (UAVs) in urban environments at different altitudes for future mobile communication systems and proposed a new ensemble learning model consisting of five base learners. In a study based on path loss data obtained from Greece's rural regions, (Moraitis et al., 2022) evaluated the performance of diverse ensemble learning models in quantifying propagation loss in such areas. In this study, a stacking technique consisting of five base learners and DNN customized as a meta-learner has demonstrated the highest level of performance. A review of the current literature shows that a limited number of academic studies have focused on two-level stacking-based ensemble models that integrate machine learning models based on augmentation to estimate signal path loss in the 1800 MHz frequency band in urban areas by integrating them with Geographic Information Systems (GIS).

This study demonstrates that the integration of ML-based radio propagation models and urban three-dimensional (3D) models offers significant contributions to the optimisation of urban communication networks. The objectives of this study are as follows: (i) generate signal coverage maps through spatial analyses; (ii) simulate path loss models using mobile station measurements, building, and Digital Elevation Model; (iii) develop and perform evaluation of binary hybrid models consisting of different boosting-based ML methods. This study provides a comprehensive assessment by integrating Geographic Information Systems (GIS) based spatial analyses with signal path loss models to assess the signal loss for each base station. A Digital Twin of the study area, has been created by combining 3D building models and Digital Elevation Model (DEM). In addition, this structure has made it possible to integrate signal path loss models into smart city applications. Two hybrid models consisting of ensemble boosting models (XGBoost, CatBoost, Gradient Boosting Machine) that take into account the effects of distance, 3D spatial data, land use/cover, and topography on signal strength are comparatively evaluated. The signal propagation and coverage area of the base stations are simulated with digital twin and GIS-supported spatial modelling and visualisation techniques.

The organization of the paper is as outlined below: Section II presents the study area and data preparation; Section III describes GIS-integrated spatial analyses and machine learning models used for estimating path loss; Section IV provides the findings and discussion; and Section V concludes the paper with the main results.

2. Study Area, Measurement Method and Data Organization

Istanbul is the world's 23rd-largest metropolis. It has a population of 16 million. The total area of Istanbul is 5,460 square kilometres. The study area has been identified as Istanbul. The study area's elevation varies from 0 to 536 meters. Downlink signal measurements have been performed in Kadıköy, Fatih and Zeytinburnu in accordance with the national frequency plan in order to train and validate the boosting based models. To implement the machine learning model, a total of three distinct categories of datasets have been incorporated into the model as inputs. Details on the study area are presented in Figure 1. Three differently structured datasets have been used as input to implement the machine learning model. The dataset consists of base station site topology, geographical data and measurement station data. The dataset has been incorporated into a geographical database. The World Geodetic System (WGS 1984), a datum that is widely known and used worldwide, is used for all our data layers.

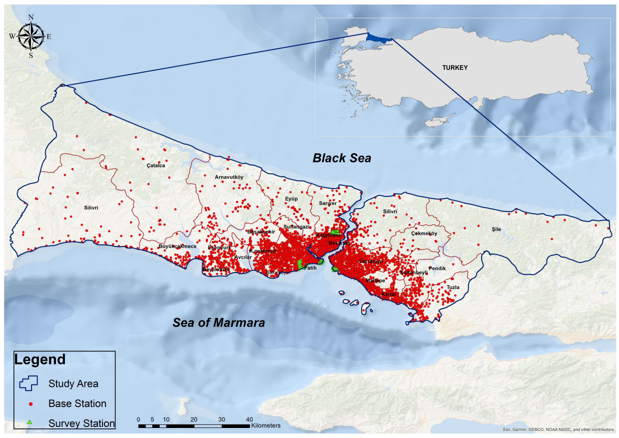

Figure 1. Study Area.

The site topology database contains the location, height, antenna power, and frequency information of the base stations in the study area. The geographical database comprises three datasets: (i) Digital Elevation Model (DEM) data, which provides land elevation above sea level at a 25 × 25 meters resolution in raster format; (ii) CORINE (Coordination of Information on the Environment) infrastructure, which classifies spatial regions into 44 different categories based on land use and land cover characteristics; and (iii) building height data, a vector polygon dataset that includes the footprints, corner coordinates, and height information of buildings within the study area.

The survey station data include the coordinates of mobile receiver base stations within the study area, the distance between the mobile station and the base station, the height of the receiver, and the Received Signal Strength (RSS) values collected in the field using the HF-6065 spectrum analyzer. Figure 2 shows the Aaronia HF 6065 portable spectrum analyzer. The 1800 MHz frequency band is based on the Turkish National Frequency Plan, and detailed band allocation information has been obtained in accordance with this plan. In the course of this study, downlink frequency division duplex (FDD) signal measurements have been conducted within the 1800 MHz band (1805.1–1879.9 MHz)(GSM & IMT-2000/UMTS & IMT Detailed Band Plan and Allocation Information Turkey, 2023).

Figure 2. Portable Spectrum Analyzer.

3. GIS Integration of Spatial Analysis for Signal Path Loss

Proximity analyses assess the spatial relationships of features (such as points, lines, polygons, or raster cells) by measuring their distances to surrounding features or cells. Proximity analysis involves calculating the distance between input features and the nearest feature in a different layer or feature class, as well as deriving additional information about their spatial neighborhood (Ness and Brogaard, 2008). In this study, the distances of mobile stations to the nearest base station are determined by proximity analysis.

Visibility analysis has been applied in the study as another spatial analysis. Visibility analysis is a type of analysis used in GIS applications to examine whether there is a direct line of sight (LOS analysis) from a specific point in an area to another point, and how much of the surrounding area can be seen from a specific observation point (Viewshed analysis) (Zhao et al., 2024). Furthermore, obstacles such as terrain height and buildings have been automatically detected in the 3D visibility analysis. In this scope, using DEM and a single observation point, a new raster image showing the visible and invisible areas for each BTS (coded with 1 and 0) has been produced. The mutual visibility problem between each BTS and each survey station in the study area is shown in Figure 3 based on visibility analysis with 3D spatial linearity and the intersection of the topographic profile.

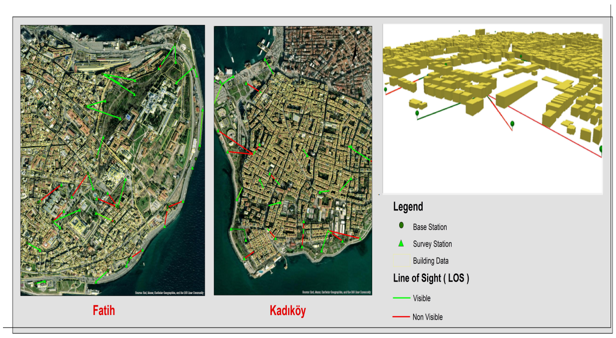

Figure 3. Visibility Analysis.

In the following stage of this study, spatial overlay analysis has been applied. In spatial analysis, the process of overlaying and merging the layers of maps belonging to different geographical features that have the same coordinate system is referred to as spatial overlay analysis. Spatial join is a spatial analysis method frequently used in GIS. Spatial join analysis is a simple way to rapidly examine spatial correlation. This method combines feature information from multiple layers within the same coordinate system by leveraging their spatial relationships (Zhang et al., 2021). The land use types where the measurement stations are located have been determined by spatial join analysis.

4. Ensemble Learning Methods for Signal Path Loss

In this study, binary hybrid models of various individual machine learning (ML) algorithms such as CatBoost, XGBoost, GBM are evaluated for a regression problem. An understanding of the individual algorithms is essential for interpreting the hybrid modeling approach; therefore, brief descriptions of each model are presented.

Gradient Boosting Machine (GBM) is a supervised learning technique capable of predicting both categorical and continuous target variables. A strong predictive model is built iteratively using multiple weak learners (typically decision trees) while minimizing the loss function via gradient descent (Aldossari, 2023).

XGBoost is an optimised and accelerated implementation of the Gradient Boosting Machine (GBM), offering superior performance especially when applied to structured datasets. Additionally, XGBoost can operate efficiently in high-dimensional feature spaces, maintaining relatively low computational cost compared to other gradient boosting algorithms (Liu et al., 2022). Assume a data set defined as $D=\{(x_i,y_i)\}$, $i = 0, 1, ..., n$ where each x_i consists of m attributes $x_i = (x_{i1},..., x_{im})$. Here, $\hat{y}_i$ represents the predicted output value for the i-th data point.

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i), \quad f_t \in F \tag{1}$$

Here, F denotes a set of regression trees; x_i represents the explanatory variables for the ith data point; the expression $f_t(x_i)$ refers to the prediction made by the tth tree in the ensemble. The final prediction is obtained by summing the outputs of all T trees, where T denotes the total number of trees used in the model (Nagao and Hayashi, 2020).

CatBoost is a type of GBM model that can automatically convert categorical features in data sets into numerical data, thereby reducing the need for data preprocessing. In complicated data sets containing categorical data, this model can process the data directly, thereby providing simpler but more effective performance. Symmetric trees are a decision tree model that divides into subtrees by splitting at each node and form the basis of the model. These symmetric trees are structurally similar to each other in that they apply the same splitting condition to all leaf nodes at each level of the tree. This eliminates the problem of overfitting and provides higher prediction accuracy without creating deep trees (Masood et al., 2023). The CatBoost model can be defined as in equation 2:

$$Z = H(x_i) = \sum\nolimits_{j=1}^{J} c_j 1_{\{x \in R_j\}} \tag{2}$$

Here, $H(x_i)$ denotes the function of a decision tree constructed based on the explanatory variables x_i. Each tree splits the feature space into disjoint regions R_j, corresponding to its leaves, based on binary splits of both numerical and encoded categorical features.

This hybrid approach utilises an ensemble learning technique known as stacking and it has been observed that this technique provides higher accuracy than other machine learning algorithms. This approach seeks to enhance prediction accuracy by leveraging the strengths of multiple basic regressors. The essence of the stacking technique lies in its hierarchical structure: This method consists of multiple basic regressors, each of which is independently trained on the dataset. The basic regressors generate predictions, which the higher-level meta-regressor then uses as input features. This meta-regressor, often termed the final estimator, combines the outputs of the base regressors to generate the final prediction. The flow chart of the aforementioned method is provided in Figure 4.

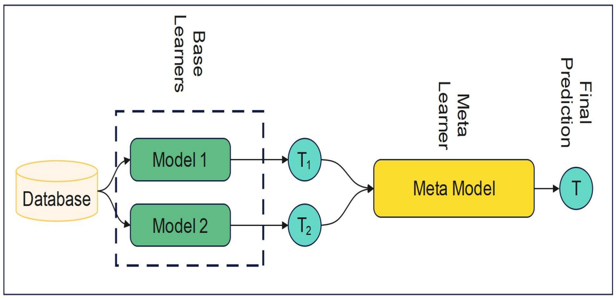

Figure 4. Flow chart of the stacking model.

ML Models	Optimum Hyperparameter	
	Hyperparameter	Optimal value
XGBoost	colsample_bytree	0.8
	learning_rate	0.1
	max_depth	8
	n_estimators	300
CatBoost	iterations	500
	learning_rate	0.1
	depth	5
	l2_leaf_reg	6
GBM	learning_rate	0.1
	max_depth	5
	n_estimators	500

Table 1. Hyperparameter Tuning Outcomes.

In this study, the most appropriate hyperparameter selection for each model has been performed using the GridSearchCV methodology in the Scikit-Learn library in Python software. Hyperparameter tuning has been performed by cross-validation. Cross-validation is a type of statistics-based resampling that objectively and accurately analyzes the efficiency of machine learning models across a variety of data sets.

$K-$ fold cross validation is employed not only to assess model performance but also to perform hyperparameter tuning to optimise generalisation capability. This procedure plays an important role in determining the stability of a machine learning model. The $K-$ fold cross validation method involves the division of the data set into K subsets. The model is subsequently subjected to training and testing K times, with each subset utilized as the test data set. The application steps and operational logic of the method at issue are presented in Figure 5 in order to illustrate the methodology more clearly.

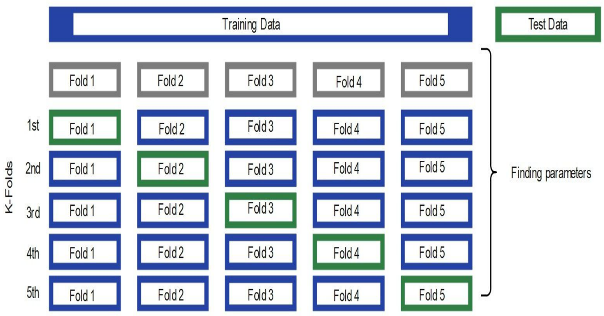

Figure 5. $K-$ fold Cross-validation Method.

Within the framework of this study, the value of k has been designated as 10 by the K-fold cross-validation approach. Hyperparameters determined for the models are given in Table I.

5. Results and Discussions

In this study, frequency, received signal strength, land use/land cover, height, base stations, and building heights collected from the field have been used for modelling. Model development has been implemented with Python JupyterLab and spatial analysis and mapping with ArcGIS. The data set consists of both categorical and numerical data types. Categorical data has been transformed with One Hot Encoding; data set has been divided into 80% for training and 20% for testing. For the model, a new data set has been created with random points to generate synthetic data. All spatial analyses have been repeated to create a new data set. Model evaluation metrics used to assess the accuracy and generalisation capacity of the ensemble method form the basis of performance analysis in regression-based problems. This study employs two evaluation metrics: the root mean squared error (RMSE) and the mean absolute error (MAE). These metrics used in this study are considered standard benchmarking criteria that allow comparison with similar algorithms. These performance metrics can be defined mathematically as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(PL_i^{Measured} - PL_i^{Pred}\right)^2} \tag{3}$$

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|PL_i^{Measured} - PL_i^{Pred}\right| \tag{4}$$

Here, $PL_i^{Measured}$ represents the signal path loss values obtained as a result of the measurements, PL_i^{Pred} represents the predicted signal path loss values, i represents the index of the measured samples, and N represents the total number of samples (Al-Thaedan et al., 2024).

In Table II, MAE and RMSE results for 1800 MHz path loss estimation have been presented. A comparative evaluation of ensemble learning approaches for path loss prediction shows that combining CatBoost and XGBoost yields superior predictive performance compared to using XGBoost and GBM together.

Specifically, the CatBoost+XGBoost model achieved a substantially lower mean absolute error (MAE) of 0.9802 compared to 1.3496 for the XGBoost+GBM model, indicating that the model produces predictions that, on average, deviate less from the true path loss values. The Root Mean Square Error (RMSE) results reinforce this finding: CatBoost+XGBoost achieved an RMSE of 1.1910, substantially outperforming the other ensemble (RMSE = 1.9767). As for the recommended stacking ensemble models, CatBoost and XGBoost stacking models perform better because CatBoost efficiently processes categorical data without requiring transformation. These results suggest that the inclusion of CatBoost, which is particularly effective in handling categorical features and mitigating overfitting through ordered boosting, contributed positively to the overall model robustness and generalization capability. By contrast, the XGBoost+GBM ensemble exhibited higher error metrics, which may reflect less effective handling of non-linear dependencies and potential redundancy between the two gradient boosting methods. Overall, the CatBoost+XGBoost ensemble offers a more accurate and reliable predictive framework for modeling path loss in the studied propagation environment.

Ensemble Models	Performance Metrics	
	MAE	RMSE
CatBoost+XGBoost	0.9802	1.1910
XGBoost+GBM	1.3496	1.9767

Table 2. Performance Metrics of Evaluated Machine Learning Algorithms.

In comparison with the results of other studies in the literature, (Sani et al., 2022) has examined two separate data sets and calculated signal path losses at multiple frequencies and in multiple environments. The study has applied Gradient Boosting and XGBoost models and obtained RMSE values between 2.75 and 5.48 dB in various 1800 MHz environments.In the study (Liu et al., 2022), the signal strength at a frequency of 3.5 GHz has been analyzed using the XGBoost model, and geographical information has been integrated into the analysis. The results have been compared with empirical models such as SUI and ECC. The XGBoost model shows RMSE values between 7 and 8 dB under different circumstances. The CatBoost+XGBoost stacking model has shown the best performance with 1.19 dB RMSE, efficiently processing categorical data. Based on previous research, this study combines GBM, CatBoost, and XGBoost models using a stacking method and demonstrates better performance with lower RMSE values. Table II also highlights the effectiveness of ensemble learning techniques generated using the stacking method in improving path loss prediction accuracy. The signal path loss estimation map covering the whole region with the Kriging method has been given in Figure 6.

According to the signal path loss map shown in Figure 6, which is generated with synthetic data, it is seen that in the districts with dense urbanisation, high building height and number, high population density significantly increase the signal loss (in the red regions of the map). In addition, other urbanised structures (e.g. bridges, tunnels) may interfere with the effective transmission of the signal.

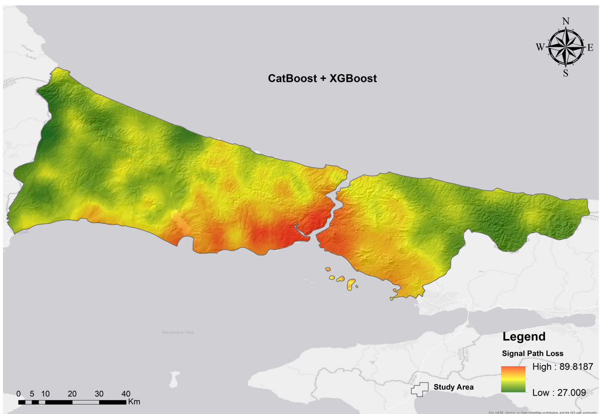

Figure 6. 1800 MHz Signal Path Loss.

6. Conclusions

Path loss models are developed to satisfy specific network requirements. The selection of a suitable path loss model for the target area is of significant importance in terms of evaluating coverage quality. In this context, Geographic Information Systems (GIS) assist in determining spatial variables for coverage estimation models and play a critical role in network propagation evaluation.

In this study, signal loss has been visualised by Kriging interpolation using Geographic Information Systems (GIS) for path loss estimation and a comprehensive loss surface has been created for the whole area. A stacking ensemble-based method, integrating a dual combination of three community learning techniques, is proposed. The findings indicate that the proposed architecture possesses the capacity to accurately identify and predict a significant portion of the measured path loss data, while effectively minimizing the mean square error. These findings are based on a test conducted on 20% of the measured dataset. Both the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE) have achieved a significant reduction, indicating improved predictive accuracy and enhanced generalization performance. Stacking ensemble learning methods have improved accuracy by combining the predictions of different models; in particular, CatBoost's direct processing of categorical data has led to high performance in analyses involving geographic information. The proposed 2D and 3D based approach has achieved successful results in terms of accuracy, efficiency, and scalability, and may make significant contributions to 6G and beyond 6G network planning.

References

Aggarwal, D., R, S. C. T., Mittal, S., 2024. A stacking ensemble technique to predict speed and distance in 4g and 5g communication datasets. Technical report.

Al-Thaedan, A., Shakir, Z., Mjhool, A. Y., Alsabah, R., Al-Sabbagh, A., Nembhard, F., Salah, M., 2024. A machine learning framework for predicting downlink throughput in 4G-LTE/5G cellular networks. *International Journal of Information Technology (Singapore)*, 16, 651-657.

Aldossari, S. A., 2023. Predicting Path Loss of an Indoor Environment Using Artificial Intelligence in the 28-GHz Band. *Electronics (Switzerland)*, 12.

Cui, Y., Yuan, W., Zhang, Z., Mu, J., Li, X., 2023. On the Physical Layer of Digital Twin: An Integrated Sensing and Communications Perspective. *IEEE Journal on Selected Areas in Communications*, 41, 3474-3490.

Fauzi, M. F. A., Nordin, R., Abdullah, N. F., Alobaidy, H. A., 2022. Mobile Network Coverage Prediction Based on Supervised Machine Learning Algorithms. *IEEE Access*, 10, 55782-55793.

Goudos, S. K., Athanasiadou, G., 2019. Application of an Ensemble Method to UAV Power Modeling for Cellular Communications. *IEEE Antennas and Wireless Propagation Letters*, 18, 2340-2344.

GSM & IMT-2000/UMTS & IMT Detailed Band Plan and Allocation Information Turkey, 2023.

Ibhaze, A. E., Atayero, A. A.-A., Ajose, S. O., Idachaba, F. E., 2016. Developing smart cities through optimal wireless mobile network. *IEEE International Conference on Emerging Technologies and Innovative Business Practices for the Transformation of Societies (EmergiTech)*, IEEE.

Kuruvatti, N. P., Habibi, M. A., Partani, S., Han, B., Fellan, A., Schotten, H. D., 2022. Empowering 6G Communication Systems With Digital Twin Technology: A Comprehensive Survey. *IEEE Access*, 10, 112158-112186.

Kwon, B., Son, H., 2024. Accurate Path Loss Prediction Using a Neural Network Ensemble Method. *Sensors*, 24, 304. https://www.mdpi.com/1424-8220/24/1/304.

Liu, Y., Dong, J., Huangfu, W., Liu, J., Long, K., 2022. 3.5 GHz Outdoor Radio Signal Strength Prediction With Machine Learning Based on Low-Cost Geographic Features. *IEEE Transactions on Antennas and Propagation*, 70, 4155-4170.

Masood, U., Farooq, H., Imran, A., Abu-Dayya, A., 2023. Interpretable AI-Based Large-Scale 3D Pathloss Prediction Model for Enabling Emerging Self-Driving Networks. *IEEE Transactions on Mobile Computing*, 22, 3967-3984.

Moraitis, N., Tsipi, L., Vouyioukas, D., Gkioni, A., Louvros, S., 2021. Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz. *Wireless Networks*, 27, 4169-4188. https://link.springer.com/article/10.1007/s11276-021-02682-3.

Moraitis, N., Tsipi, L., Vouyioukas, D., Gkioni, A., Louvros, S., 2022. On the Assessment of Ensemble Models for Propagation Loss Forecasts in Rural Environments. *IEEE Wireless Communications Letters*, 11, 1097-1101.

Morocho-Cayamcela, M. E., Lee, H., Lim, W., 2019. Machine learning for 5G/B5G mobile and wireless communications: Potential, limitations, and future directions. *IEEE Access*, 7, 137184-137206.

Nagao, T., Hayashi, T., 2020. Study on radio propagation prediction by machine learning using urban structure maps. *14th European Conference on Antennas and Propagation (EuCAP).*

Ness, B., Brogaard, S., 2008. GIS proximity analysis and environmental assessment of sugar beet transport in Scania, Sweden. *Area*, 40, 459-471.

Ojo, S., Akkaya, M., Sopuru, J. C., 2022. An ensemble machine learning approach for enhanced path loss predictions for 4G LTE wireless networks. *International Journal of Communication Systems*, 35.

Sani, U. S., Malik, O. A., Lai, D. T. C., 2022. Dynamic Regressor/Ensemble Selection for a Multi-Frequency and Multi-Environment Path Loss Prediction. *Information (Switzerland)*, 13.

Shaibu, F. E., Onwuka, E. N., Salawu, N., Oyewobi, S. S., Djouani, K., Abu-Mahfouz, A. M., 2023. Performance of Path Loss Models over Mid-Band and High-Band Channels for 5G Communication Networks: A Review. *Future Internet*, 15.

Sotiroudis, S. P., Boursianis, A. D., Goudos, S. K., Siakavara, K., 2022. From Spatial Urban Site Data to Path Loss Prediction: An Ensemble Learning Approach. *IEEE Transactions on Antennas and Propagation*, 70, 6101-6105.

Xia, H., Liu, Z., Efremochkina, M., Liu, X., Lin, C., 2022. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society*, 84.

Zhang, A., Yang, Y., Chen, T., Liu, J., Hu, Y., 2021. Exploration of spatial differentiation patterns and related influencing factors for National Key Villages for rural tourism in China in the context of a rural revitalization strategy, using GIS-based overlay analysis. *Arabian journal of Geosciences*, 14, 1-15. http://bzdt.nasg.gov.cn.

Zhang, J., Gao, W., Wang, S., 2024. Research on Smart City Platform Construction Technology for Digital Twins. *International Journal of Advanced Network, Monitoring and Controls*, 9.

Zhao, J., Xu, A., Zhang, X., Zhang, Y., Zhang, Y., Cao, M., Huang, M., 2024. Overview and prospects of visibility analysis approaches. *the 31st International Conference on Geoinformatics*, MDPI AG, 28.