Mobile Network Coverage Prediction Based on Supervised Machine Learning Algorithms

MOHD FAZUWAN AHMAD FAUZI [1,2], **ROSDIADEE NORDIN** [2], **(Senior Member, IEEE),**
NOR FADZILAH ABDULLAH [2], **(Member, IEEE),**
AND HAIDER A. H. ALOBAIDY [2], **(Graduate Student Member, IEEE)**

[1] Malaysian Space Agency (MYSA), Kuala Lumpur 50480, Malaysia
[2] Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Selangor 43600, Malaysia

Corresponding author: Rosdiadee Nordin (adee@ukm.edu.my)

ABSTRACT The need for wider coverage and high-performance quality of mobile networks is critical due to the maturity of Internet penetration in today's society. One of the primary drivers of this demand is the dramatic shift toward digitalization due to the Covid-19 pandemic impact. Meanwhile, the emergence of the 5G wireless standard and the increasingly complex actual operating environment of mobile networks make the traditional prediction model less reliable. With the recent advancements and promising capabilities of machine learning (ML), it is seen as an alternative to the traditional approaches for ground to ground (G2G) mobile communication coverage prediction. In this study, various ML models have been tested and evaluated to develop an ML-based received signal strength prediction model for mobile networks. However, the challenge is to identify a practical ML model that can fulfill the computing speed criteria while still meeting the prediction accuracy. A total of six categories of ML models, namely Linear Regression (LR), Artificial Neural Network (ANN), Support Vector Machine (SVM), Regression Trees (RT), Ensembles of Trees (ET), and Gaussian Process Regression (GPR) that consists of more than 20 types of established algorithms/kernels have been tested and evaluated in this paper to identify the best contender among them, in terms of speed and accuracy. Findings from the evaluation showed that the GPR model is the most accurate model for Reference Signal Received Power (RSRP) prediction in terms of *RMSE* and R^2, followed by ET, RT, SVM, ANN and LR. Nevertheless, prediction speed and model training times are also important factors in determining the most practical model for RSRP prediction for several real-world mobile network planning applications. Finally, the ET model with Random Forest (RF) algorithm has been selected and highly recommended as the most practically employed ML model for developing rigorous RSRP predictions model in multi-frequency bands and multi-environment. The developed prediction model is capable of being utilized for the network analysis and optimization.

INDEX TERMS Machine learning, MATLAB, wireless communication, received signal strength indicator.

I. INTRODUCTION

As of June 2021, the fifth generation of wireless communication standard 5G, also known as New Radio (NR), has been deployed in 58 countries [1]. The 5G NR systems enable the world community to access and share information in various scenarios with extremely low latency and high data rates [2], [3]. From an industry point of view, the 5G NR catalyze economic growth further [4], [5]. Among the

The associate editor coordinating the review of this manuscript and approving it for publication was Young Jin Chun.

economic sectors that will benefit from the 5G network are automotive [6], media and entertainment, manufacturing, logistics, agriculture, energy and utilities, education and healthcare [7]–[9]. According to [10], the adoption of 5G could be even faster and expected to reach the one billion user mark within 3.5 years. 5G market in 2021 is estimated at USD 46.61 billion and is forecasted to grow to approximately USD 664.75 billion in 2028 [11].

However, it is undeniable that the deployment of 5G communication networks is a challenging task [12]. Based on the standards categorized by 3GPP (3rd Generation

Partnership Project), 5G is divided into two categories, namely Sub-6GHz 5G, which operates in the same spectrum band as 4G LTE (450 MHz to 6 GHz), while mmWave 5G operates in 24.25 GHz to 52.6 GHz spectrum band [13]. For 5G networks operating at mmWave, the signal itself is more vulnerable to the effects of the transmission environment, which consequently shortens the effective communication distance [3], [5], [6]. The antennas need to be placed nearer to users, which translates to about ten times more antennas to be installed compared to the previous 4G requirement (approximately 100 to 200 meters apart) [2], [5], [6], [14]. Therefore, 5G network development requires extensive and critical radio network planning and analysis. It will involve numerous types of data and factors to identify the best positions for antenna deployment [15]. Because of the complexity of planning and analysis works which are not feasible to be executed using traditional methods, a high-level data analytics adoption using machine learning (ML) is required [5], [9]. This is coherent with the latest technical specifications issued by 3GPP (Release 18), where ML capabilities are beginning to be adopted as part of advanced network planning for 5G future deployments [16].

ML is a branch of artificial intelligence (AI) and computer science that focuses on the use of data and algorithms and has recently gained popularity in the field of wireless communication[2], [17], [18]. The ML-based prediction model is seen as a game-changer in modern mobile network planning due to its capability to produce more accurate results than the traditional empirical-based prediction model. It is also more efficient in terms of data processing capability compared to the deterministic-based prediction model [12], [19]–[22]. Besides, the ML-based models can improve their accuracy over time without having to be programmed [9].

In traditional prediction methods, mobile network planning is inflexible [22]. The prediction is constrained to certain specifications and conditions such as frequency, antenna height and environmental characteristics. However, the reality is that the operating environment of modern radio networks is more diverse and complex [23]. Therefore, it is necessary to explore and build a new prediction method that can operate more flexibly and universally to adapt to modern mobile networks' operation complexities. For ground-to-ground (G2G) mobile communication scenarios, the authors in [18], [22], [24]–[28] have tested and evaluated the capabilities of several ML models for the mobile networks' performance prediction. Overall, the results of the studies proved that the ML-based prediction model had outperformed the traditional methods in terms of computational efficiency, applicability, and accuracy.

ML-based prediction models for path loss in an urban environment have been implemented in Beijing, China, in which authors in [27] using Artificial Neural Network (ANN), Support Vector Regression (SVR) and Random Forest (RF) models. The root-mean-square error (**RMSE**) value generated is in the range of 4 dB to 5 dB. The work was performed on

the frequency spectrum of 877.26 MHz, 2021.4 MHz, and 2127.6 MHz with the input parameters used are: (i) the separation distance between transmitter (T_x) and receiver (R_x); and (ii) frequency. The work was based on 2,558 datasets with ratios of 53.6% and 46.4% for model training and testing, respectively. Similarly, in [25], the work was conducted in an urban environment in Lisbon, Portugal, at 3.7 GHz and 26 GHz frequency bands, using a real 5G network. The input parameters and response variables used are almost the same as [27]. However, the used dataset is two times larger. Meanwhile, the models studied are only SVR and RF, with **RMSE** values of around 6 dB to 7 dB.

Studies for the urban environment using a simulation dataset were carried out by [18] on the 2140 MHz frequency band only. A total of 5,150 datasets were generated using ray-tracing techniques. Like the previous two studies, the response variable is path loss. However, three new input parameters were introduced in this work, i.e.: (i) T_x height and R_x height; (ii) coordinates of R_x; and (iii) the status of signal propagation path, being either line of sight (LOS) or non-line of sight (NLOS). The predictive results are based on SVR and RF models with **RMSE** value between 2 dB to 4 dB. However, according to authors in [26], the R_x location ought to be eliminated to get a prediction model with better generalization and adaptability that could be advantageously employed in various geological areas.

The use of simulated datasets on urban environments has also been carried out by [24], but the focus is only on NB-IoT networks, a branch of mobile communication that uses low bandwidth. The frequency spectrum tested were at 900 MHz and 1800 MHz with the input parameters used are: (i) the separation distance between T_x and R_x; (ii) the height of the building involved; and (iii) the signal propagation path status (LOS/NLOS). The used dataset was almost seven times larger than [18], with sampling ratios of 80% and 20% for training and testing, respectively. However, only ANN and RF models were tested and evaluated with **RMSE** values around 4 dB to 6 dB.

It has been implemented by [28] for suburban environments on the 450 MHz, 1450 MHz, and 2300 MHz frequency bands in South Korea. As in previous studies, the input parameters used are almost the same except for introducing a new input parameter, which is the ratio of T_x height to the R_x height. However, the authors did not clearly state the actual size of the dataset. The ML models studied were only ANN and GPR, with *RMSE* value between 8 dB to 9dB for both models.

In [26], the ML-based prediction model was tested in a rural environment in Greece using ANN, SVR, and RF models. The **RMSE** value is between 4 dB to 5 dB. This work was only implemented on a single frequency band of 3.7 GHz. The input parameters used are as follows: (i) 3-dimensional (3D) distance between T_x and R_x; (ii) T_x height and R_x height at the above sea level (ASL); and (iii) the signal propagation path status (LOS/NLOS). Meanwhile, the response variable is

path loss. The dataset was around 2,200, with a sampling ratio of 70%, 15%, and 15% for training, validating, and testing, respectively.

Lastly, the research work in [22] is more comprehensive than the other studies mentioned earlier. The ML-based model has been tested in multi-environments, a combination of dense cities, open areas, and inland lakes. The input parameter used is also quite comprehensive, consisting of (i) 2-dimensional (2D) and 3D separation distance between T_x and R_x; (ii) frequency; (iii) height difference between T_x and R_x; (iv) T_x tilt angle; (v) T_x azimuth angle; (vi) transmitting power; (vii) clutter and building heights information; and (viii) the vertical distance of the R_x from the major lobe signal line of the T_x antenna. Besides, this is the only one work that uses Reference Signal Received Power (RSRP) as the response variable. In addition, the dataset is very large, consisting of 12,011,833 datasets. Unfortunately, it is only tested on an RF model with an **RMSE** value of 6.11 dB. The works mentioned above are further summarized in Table 1 for a clear comparison.

Although there is no specific rule of thumb for determining the number of datasets suitable for developing an ML-based mobile network performance prediction model, we found that a larger dataset may provide more consistent predictive performance even if the dataset is generated through computer simulation. In this study, the response variable for the developed prediction model is RSRP because it is the key parameter directly representing the state of network signal level at the UE location [29] in 4G LTE and 5G NR networks. We decided to use RSRP instead of path loss in this study to manifest the final output in a more meaningful way. In the meantime, the ML models tested in the previous works for G2G mobile communications are only ANN, SVR, RF, and GPR. Besides, the applied algorithms/kernels of the models are limited to a certain type only.

AS Such, the contributions of this study can be summarized as follows

- An extensive measurement campaign consists of 21,323 datasets collected in various outdoor environments using an Android-based drive test solution, which the cleaned dataset is available for download at [31].
- A guide to generating input parameters using online-based radio planning tools (CloudRF)
- A comprehensive methodology is presented in preparing a clean study dataset for ml-based rsrp predictive modeling development
- Assessment and validation of different state-of-the-art ML models and their performance for the application in G2G mobile networks signal strength prediction.
- Development of ml-based model for faster mobile network coverage prediction
- The ML-based RSRP prediction model is developed for various deployment scenarios, such as in urban, suburban, open areas, etc. In the meantime, the model is also compatible with multiple frequency bands.

- Evaluation and optimization approach for received signal strength in existing 4g lte mobile networks
- Performance analysis of ML-based model against previous work done by [22], [25], but considering an extension of scenarios and other improvements in methodology, as shown in Table 1.

The rest of the paper is organized as follows: Section II describes the principle of the considered ML model. Section III presents the experimental setup for the data collection and datasets preparation. The learning, testing, and validation of ML models are outlined in Section IV. The performance of all tested ML models is evaluated and discussed in Section V, followed by model applications and optimization approaches in Section VI. Finally, Section VI concludes the article.

II. MACHINE LEARNING BASED MODELS FOR MOBILE NETWORKS RSRP PREDICTION

Predicting the RSRP of mobile networks can be categorized as a regression type problem [26], [32]. Regression problems are a part of supervised ML techniques, where the models are trained based on a given labeled data [27]. The state-of-the-art supervised ML-based models that will be tested and examined in this study are discussed below.

A. LINEAR REGRESSION (LR)

LR is an ML-based method that finds the linear relationship between the input parameters and response variable [33], [34]. LR is easy to implement and easy to interpret the results. However, its disadvantage is the initial assumption of a linear relationship between input parameters and response variables, which is not suitable when dealing with non-linear relationships [35].

B. ARTIFICIAL NEURAL NETWORK (ANN)

ANN is a network of artificial neurons that mimic the way of human brain functional [36]. The basis of the ANN operation is to find the best correlation to represent the relationship between the input parameters and the response variable [26]. ANN is suitable for non-linear regression problems and shows good predictive performance with a large dataset [27]. ANN is built by a network of three segments: the input layer, the hidden layer, and the output layer, where data processing and forecasting activities are carried out on the hidden layer [36]. Therefore, the ANN prediction performance is highly influenced by the settings on the hidden layer. However, excessive use of hidden layers will cause the models to become more complex and tend to discourage generalization capabilities [37], [38].

C. SUPPORT VECTOR MACHINE (SVM)

SVM, also known as SVR, is a supervised ML method using kernel functions to solve regression problems. The kernel converts the dataset to different dimensions to obtain the best

TABLE 1. Related work for ml-based models in predicting mobile communication networks performance.

Ref	Frequency (MHz)	Environment	Input Parameters	Response Variable	Training Method	Tested Models	RMSE (dB)
[27]	877.26 2021.4 2127.6	Urban (Beijing, China)	• Distance T_x to R_x • Frequency	Path Loss	2,558 datasets (53.6% Training & 46.4% Testing)	ANN SVR RF	4.74 4.54 4.19
[25]	3700 26000	Urban (Lisbon, Portugal)	• 3D distance T_x to R_x • Frequency	Path Loss	6,306 datasets (10-fold Cross-Validation)	SVR RF	6.97 6.25
[18]	2140	Simulation Urban (Frankfurt)	• Distance T_x to R_x • Frequency • Height of T_x • Height of R_x • Coordinate of R_x • LOS/NLOS	Path Loss	5,150 datasets (10-fold Cross-Validation)	SVR RF	$2.2 - 4.1$ $2.1 - 3.6$
[24]	900 1800	Simulation Urban	• Distance T_x to R_x • Height of building • LOS/NLOS	Path Loss	34,501 datasets (80% Training & 20% Testing)	ANN RF	$4.39 - 5.61$ $4.38 - 5.60$
[28]	450 1450 2300	Suburban (Korea)	• Distance T_x to R_x • Frequency • Height of T_x • Height of R_x • Height ratio T_x to R_x	Path Loss	5-fold Cross-Validation	ANN GPR	8.52 8.94
[26]	3700	Rural (Greece)	• 3D distance T_x to R_x • Height of T_x (ASL) • Height of R_x (ASL) • LOS/NLOS	Path Loss	~2,200 datasets (70% Training, 15% Validating & 15% Testing)	ANN SVR RF	4.0 4.4 4.3
[22]	2500	Dense cities, open areas & inland lakes (China)	• 2D and 3D distance T_x to R_x • Frequency • Height difference T_x to R_x • T_x tilt angle • T_x azimuth angle • Transmit power • Clutter & building heights • Vertical distance of R_x from the major lobe signal line of the T_x antenna	RSRP	12,011,833 datasets (10-fold Cross-Validation)	RF	6.11
This Study	1800 2600	Urban, Suburban, Open Areas (Putrajaya, Malaysia)	• 3D distance T_x to R_x • Frequency • Height ratio T_x to R_x • Tilt angle of transmitted signal to R_x • LOS/NLOS	RSRP	18,048 datasets (10-fold Cross-Validation)	ANN SVR RF GPR RT LR	6.82 6.62 6.18 5.64 6.46 8.65

hyperplane settings to represent the correlation between input parameters and response variables [18], [26]. Thus, non-linear dataset can be map into linear relationships at higher dimensional spaces [27], [32], [39]. SVM is less prone to overfitting and has excellent generalization capabilities but is less efficient when dealing with large datasets, particularly with a lot of noise [35], [40].

D. REGRESSION TREES (RT)
RT is a decision tree-based method that uses a tree-like structure in making predictions based on the rules set at each node [23], [41]. Thus, RT performance is highly influenced by the number of nodes from the root to the leaf [42]. Data processing and analysis in RT are easy to interpret. It is also capable of handling issues related to missing values in the

study datasets; However, it is easily affected by noise, leading to overfitting if the tree setting is too deep [35], [40].

E. ENSEMBLES OF TREES (ET)

ET is an approach of combining multiple RT to produce an output [24]. In principle, ET is a group of weak learners combined to form a strong learner [43]. Therefore, it requires more computational power and a longer training time than RT. There are two types of ET techniques: Bagging or Boosting [44]. Bagging is a technique to decrease the variance in the prediction by generating subsets of data chosen randomly with replacement from the original training dataset [26]. Meanwhiles, Boosting is an iterative technique that adjusts the weight of an observation based on the prior learner's result [15]. ET can overcome the overfitting issues in RT [35], [45], which is more robust to the noise. RF is one of the popular ET algorithms that uses bagging techniques in performing regression predictions [26], [46].

F. GAUSSIAN PROCESS REGRESSION (GPR)

GPR is a Bayesian non-parametric model [15] that utilizes kernel functions in solving regression problems [47]. It derives the relationship between input parameters and response variables from unknown functions [28]. GPR can produce good predictions at high dimensions, even with small datasets [48], and support non-linear and complex problems [49]. However, the drawback of GPR is that it requires high computational power [50], [51].

III. DATA COLLECTION AND DATASET PREPARATION

Fig.1 describes the overall concept of execution of this study. The study dataset is constructed from a comprehensive measurement campaign conducted in the Federal Territory of Putrajaya, a planned city that serves as the administrative capital of Malaysia. The selection of Putrajaya was based on the following factors: (i) has a unique town planning architecture that combination of multi-environment characteristics, i.e., high-rise and mid-rise buildings, single and double story terrace houses, lakes, densely-vegetated parks and open areas in one territory [6], [52] as shown in Fig. 2 and (ii) one of the testbed location for 5G NR network deployment beside Cyberjaya and others several major cities [53], [54]. However, at the time of this study, there is still no 5G NR mobile network officially operating for public consumers. Therefore, RSRP reading measurement can only be performed on the existing 4G LTE networks. Even though RSRP in 4G LTE and 5G NR networks have different characteristics, but the purpose of its use is the same, which is UE periodically measures RSRP for performing cell selection/reselection and handover process [30]. Since the development of the prediction model is based on a modular approach, so it can be easily extended for 5G NR network parameters in the future.

The hardware and software involved in the measurement campaign are highlighted in Table 2, while Fig. 3 shows the data type and relationship between the platforms in the measurement. To minimize the fast fading effect especially
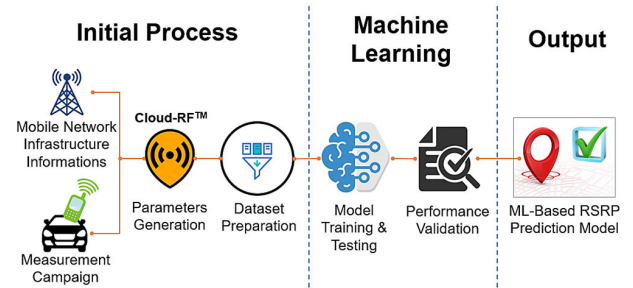

FIGURE 1. Workflow of the ML-based model for signal strength prediction adopted in this study.
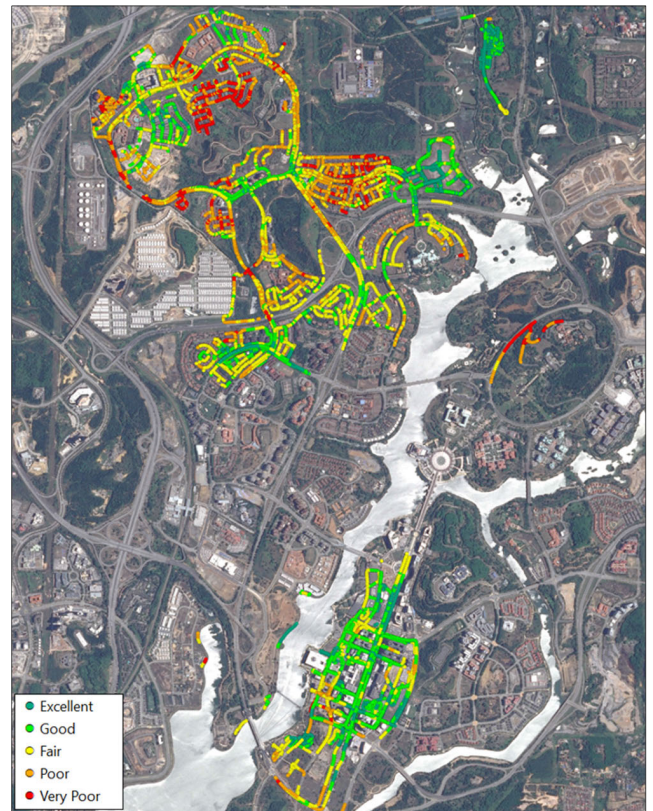

FIGURE 2. Distribution of RSRP measurement points in putrajaya.

due to the Doppler shift, the measurement campaign was conducted at a vehicle speed below 40 km/h [21].

Before dataset preparation activities are executed, the drive test data needs to be cleaned. The data collected at static conditions need to be removed to ensure the data are free from any errors [57]. After the data cleaning process was completed, UE location information in decimal degree format was extracted into the .csv file format. It was used as a reference input for the generation of the following parameters: (i) 2D distance between eNodeB (eNB) and UE; (ii) height of eNB antenna at above sea level (ASL); (iii) height of UE at ASL; angle of signal from eNB to the position of UE (tilt); and (iv) signal propagation path status between eNB

TABLE 2. Parameters of measurement campaign setup.

HARDWARE:	
Smartphone	Realme XT (Model RMX1921)
Consumer Equipment (UE)	Category 12
3GPP Release	Rel 11 (also known as LTE-Advanced)
SOFTWARE:	
G-NetTrack Pro [55]	A non-rooted wireless network monitor and drive test tool
4G LTE Switcher [56]	An android application forced the phone only to operate on 4G LTE

FIGURE 3. Description of the platform used in the measurement campaign.

FIGURE 4. Cross-section of the measurement setup.

TABLE 3. List of input parameters and explanations.

Input Parameter	Definition	Purpose
D_3	3D distance between eNB antenna and UE	To get an effective signal travel distance between T_x and R_x
H_R	Height ratio of eNB antenna to the height of EU at ASL	To get an effective height ratio between T_x and R_x
Tilt	The angle of transmitting the signal at the horizon axis from the eNB antenna to the UE	To determine the tilt angle of signal travel from the eNB antenna to the location of R_x
Fq	The use of frequency bands 1800 MHz and 2600 MHz, which are dedicated 4G LTE spectrum in Malaysia allocated by the Malaysian Communications and Multimedia Commission (MCMC)	To get a real RSRP reading in multiband measurement in Sub-6GHz frequency
Obs	The status of signal propagation path between eNB and UE	To class the signal status either LOS or NLOS

and UE. These parameters are created utilizing a web-based radio planning tool called CloudRF [58]. The terrain and clutter reference data in CloudRF are 10 meters in resolution. Besides, CloudRF also integrate high-resolution 3D buildings information from OpenStreetMap [59].
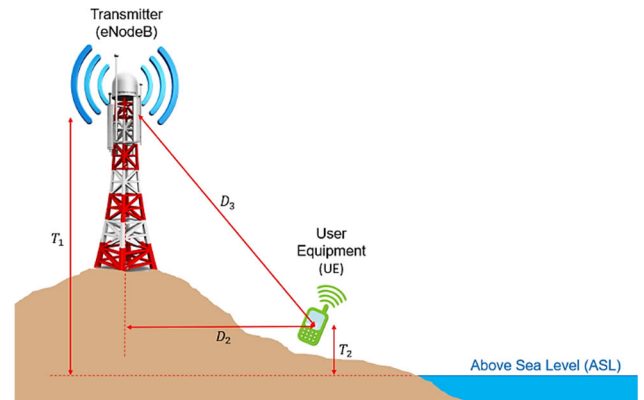
As shown in (1) and (2), two additional parameters were derived from the above-generated parameters, i.e., the 3D separation distance between eNB and UE (**D₃**)[22], [25], [26] and the height ratio between eNB antenna and EU (**H_R**) [28]. Based on [60], the concept of calculation of these two parameters applied in this study is illustrated in Fig. 4 whereas **T₁** and **T₂** are representing the height of eNB antenna and height of UE at ASL respectively, while **D₂** is the 2D separation distance between eNB and UE. Therefore, **D₃** is defined by:

$$D_3 = \sqrt{(T_1 - T_2)^2 + D_2^2} \tag{1}$$

while H_R was calculated using the following formula:

$$H_R = T_1/T_2 \tag{2}$$

Signal propagation path status between eNB and UE was labeled as '0' to represent no obstacle present (LOS); meanwhile, '1' represents NLOS condition. This parameter was labeled as Obstacle (Obs). The input parameters mentioned above are further summarized in Table 3 for a clear explanation. The selection of these five parameters as input data was based on the knowledge of electromagnetic wave propagation, which has been applied in previous works (as summarized earlier in Table 1).

The F_q and RSRP were extracted directly from the measurement campaign data. Before implementing ML model training and testing activities, outliers in D_3, H_R, Tilt, and RSRP were identified and removed. This process was done using interquartile range (IQR) analysis on 21,323 datasets, as shown in Fig. 6. Finally, a total of 18,048 cleaned datasets were prepared and ready for the ML model training and testing process. Overall, the above process is summarized and illustrated in the flow chart shown in Fig. 5.

IV. MODEL TRAINING AND VALIDATION
MATLAB 2020a Regression Learner and Neural Net Fitting application were used to train and validate the ML-based RSRP predictive model. The simulation is performed using an Intel Core i5 10th gen laptop with an onboard Radeon
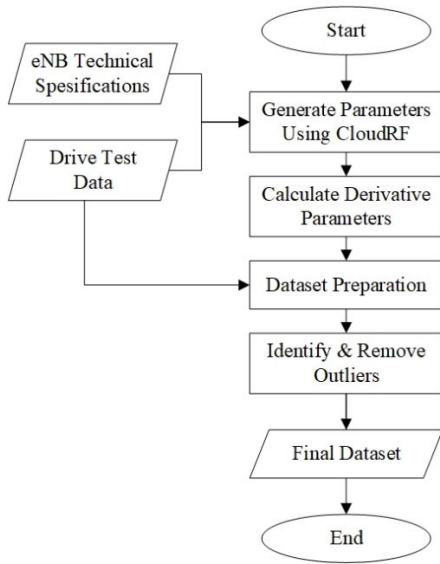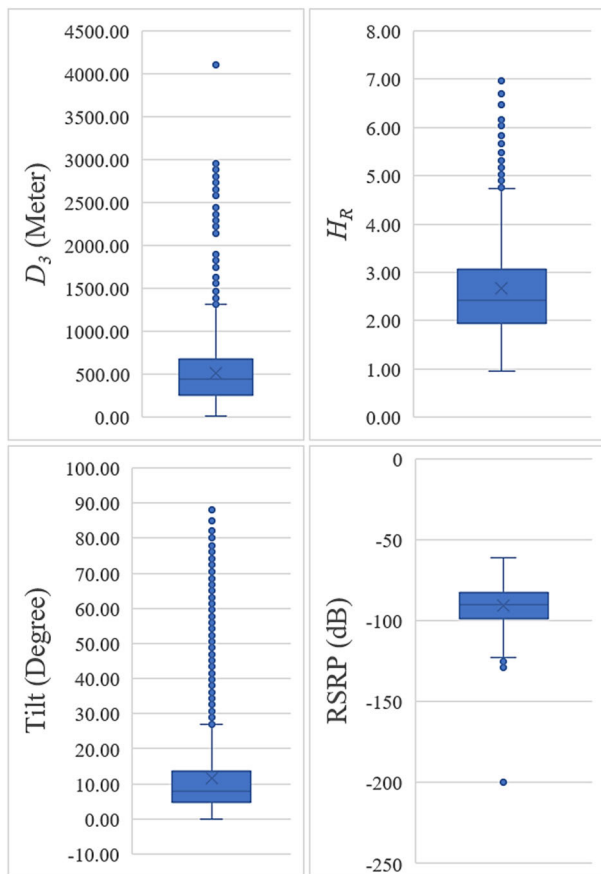
FIGURE 5. Dataset preparation process.

FIGURE 6. IQR analysis using box and whisker Plots.

530 series graphic processor, 512 GB SSD storage and 20 GB RAM.

The regression learners utilized in this study are summarized in Table 4. All these learners were examined using 10-fold cross-validation (CV), a resampling method

that split up the dataset into ten portions and then train and test procedure executed on different iterations to avert overfitting, based on a similar work done by authors in [18], [22], [25]. The parallel training function was disabled to optimize the machine processing power on each examined algorithm/kernel.

Meanwhile, the Neural Net Fitting application used a two-layer feed-forward neural network with 40 hidden neurons, as shown in Fig. 7. 70% of the dataset was used for training, and the remaining 30% was split up evenly for validation and testing purposes. Different algorithms were examined, such as Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient.

TABLE 4. List of tested regression learner.

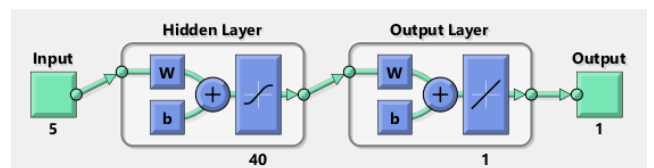| Model	Algorithm/Kernel
LR	Linear, Interactions Linear, Robust Linear, and Stepwise Linear
SVM	Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, and Coarse Gaussian
RT	Fine Tree, Medium Tree, and Coarse Tree
ET	Boosted Trees and Bagged Trees
GPR	Rational Quadratic, Squared Exponential, Matern 5/2, and Exponential

FIGURE 7. The architecture of the neural network used in this study.

To examine and validate the performance of each ML model, it is important to assess the statistical error between the measured and the predicted RSRP values. RMSE, as shown in (3), is a commonly used metric to evaluate the performance of the regression prediction models. It is given, in decibels [22], by:

$$RMSE = \sqrt{\frac{1}{n_{sample}} \sum_{i=0}^{n_{sample}-1} \left| y_i - \hat{y}_i \right|^2} \qquad (3)$$

where n_{sample} is the total number of samples, y_i is actual value, and $\hat{y}_i$ is predictive value. The smaller values of *RMSE* indicate a better prediction of the ML model. According to [26], predictive models with *RMSE* values less than 7 dB is considered acceptable, especially in an urban environment.

On the other hand, we used the coefficient of determination (R^2), as shown in (4), to reveal the degree of performance of the prediction models. It is used to describe how well the input

TABLE 5. Details performance of examined regression learner models.

Algorithm/ Kernel	RMSE (dB)	R^2	Prediction Speed (Observation/ Second)	Training Times (Second)
Model: LR				
Linear	9.01	0.36	220,000	2.00
Interactions Linear	8.65	0.41	380,000	0.58
Robust Linear	9.02	0.36	470,000	0.77
Stepwise Linear	8.65	0.41	310,000	16.49
Model: SVM				
Linear	9.03	0.36	50,000	304.85
Quadratic	8.66	0.41	36,000	1158.60
Cubic	9.00	0.37	31,000	6704.60
Fine Gaussian	6.62	0.66	13,000	126.59
Medium Gaussian	7.93	0.51	12,000	109.87
Coarse Gaussian	8.64	0.42	12,000	108.02
Model: RT				
Fine Tree	6.45	0.67	570,000	6.31
Medium Tree	6.46	0.67	880,000	1.00
Coarse Tree	6.95	0.62	930,000	0.38
Model: ET				
Boosted Trees	8.66	0.41	170,000	4.55
Bagged Trees	6.18	0.70	72,000	6.11
Model: GPR				
Rational Quadratic	5.60	0.75	2,500	1263.50
Squared Exponential	6.13	0.71	7,300	408.57
Matern 5/2	5.95	0.72	4,500	566.82
Exponential	5.64	0.75	5,100	521.40

TABLE 6. Details performance of examined ann algorithms.

Algorithm	RMSE (dB)	R^2	Epoch (At Max 1000)	Training Times (Second)
Levenberg-Marquardt	7.06	0.60	122	10.20
Bayesian Regularization	6.82	0.64	910	77.20
Scaled Conjugate Gradient	8.16	0.48	148	1.40
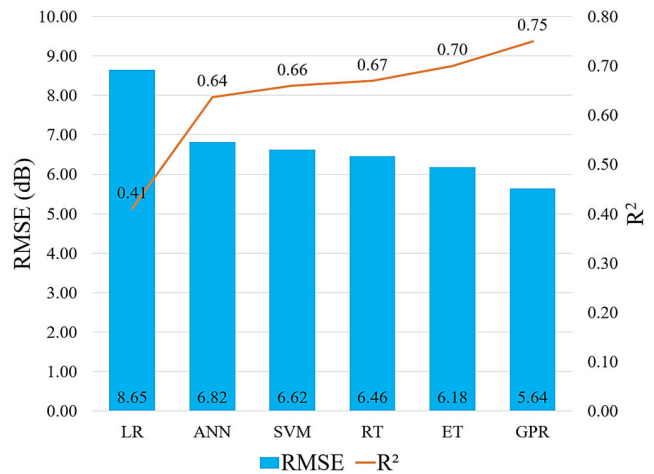
FIGURE 8. RMSE versus R^2 according to the best contender of each model category.

parameters in the model explain the response variable's variability. The larger R^2 values, the more variability is explained by the model. It is given by [28]:

$$R^2 = 1 - \frac{\sum_{n=1}^{n}(y_n - \hat{y}_n)^2}{\sum_{n=1}^{n}(y_n - \bar{y}_n)^2}, \text{ where} \bar{y} = \frac{1}{n}\sum_{n=1}^{n} y_n \quad (4)$$

Among the other important factors observed while validating the model are the duration of training times and prediction speed, which always indicate the level of complexity and efficiency of the model. Therefore, a predictive model with a balanced performance between *RMSE*, R^2, training time, and prediction speed is highly recommended as the most practically employed ML-based model for developing rigorous RSRP prognoses model in multiband and multi-environment.

V. RESULTS AND DISCUSSIONS
Details performance of examined ML models is shown in Table 5 and Table 6. Subsequently, the best contender of each

model category was selected and plotted into a Pareto chart, as shown in Fig. 8.

Based on the results, we can conclude that the GPR model outperforms others in terms of *RMSE* and R^2. This is because GPR, a non-parametric kernel-based probabilistic model, can handle small and large size datasets with fewer errors, even in high-dimensional space. GPR able to learn from the overall distribution of datasets and properly tune the hyperparameters setting to produce smoothing results. However, the GPR comes with drawbacks such as high computational power requirements and longer processing time, especially when dealing with large and imbalanced datasets. Imbalanced datasets happen when a significant disproportion within the datasets can cause unequal class distribution.

The next best performance of *RMSE* and R^2 values are achieved by the ET model with the Bagged Trees algorithm (also known as RF). RF is faster than GPR in terms of processing time. Compared to the Exponential-GPR, which is the most balanced performance model in the GPR category, RF is 85 times faster in training duration. Besides, it can also achieve around 72,000 observations per second compared to Exponential-GPR, with just around 5,100 observations per second. This is because RF, which is based on the decision tree method, is simpler to implement than GPR models.

The model category that ranked third in this study was RT with the Medium Tree algorithm has shown the most balanced performance compared to other RT algorithms. The *RMSE* and R^2 values achieved were 6.46 dB and 0.67, respectively. Although both RT and RF are based on the decision tree technique, the bagging working mechanism in RF has shown its superiority in producing better prediction accuracy than RT, which only relies on the prediction results from one decision tree only.

The fourth-placed was held by SVM. Although SVMs are kernel-based models like GPR, that can manipulate the datasets at different dimensional spaces, but the best predictive performance observed in this study is only 6.62 dB for *RMSE* and 0.66 for R^2. This is possible due to the limitations of the SVM model's capability, which is less efficient when dealing with large datasets.

The fifth place was occupied by ANN models with the best *RMSE* and R^2 results were 6.82 dB and 0.64, respectively. Although ANN accuracy is just 0.64 dB less than RF, but it is required 12.6 times longer in terms of model training times. This is due to the complex data processing analysis running in the hidden layers. Typically, to improve the accuracy, the number of hidden layers must be increased, but this will make training time and prediction speed becomes worse. In some cases, increasing the number of a hidden layer will not improve the prediction accuracy, but it could be the worst, which deters the model generalization capability. As a result, the model could not perform accurate predictions when a new input dataset was introduced. Therefore, increasing the number of hidden layers in the ANN needs to be done with caution.

Lastly, the LR models with the best result are 8.65 dB for *RMSE* and 0.41 for R^2. These results directly indicate that the LR model is unsuitable for developing the ML-based RSRP predictive model. This is possible due to the limitation of the LR-based model, which is not suitable to apply to non-linear relationships problems.

Concerning the study's objective of generating a more flexible, and fast-paced prediction model, RF is proven to be the most practical option for developing an optimal ML-based RSRP predictive model. To further enhance the RF model prediction capabilities, the optimum hyperparameter setting must be identified. To achieve this, the Optimizable Ensemble function has been executed and the result is shown in Fig. 8. The newly generated values of *RMSE* and R^2 are equal to 5.74 dB and 0.74, respectively. Therefore, RF prediction accuracy was increased by 0.44 dB and model variability has increased by 0.04. As a result, the optimum hyperparameter settings for the RF model are 495 number of learners, two minimum leaf sizes, and three number of predictors to sample.

In Fig. 9, we compare the RF model prediction performance with the previous work done by [22], [25]. The newly generated values of *RMSE* and R^2 of the RF model in this study is comparable to the values in the previous works. Although [22] uses more diverse input parameters with
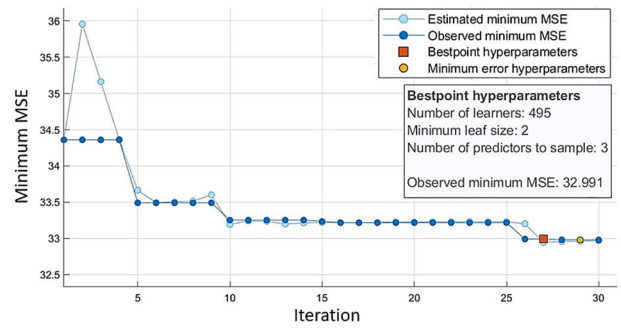

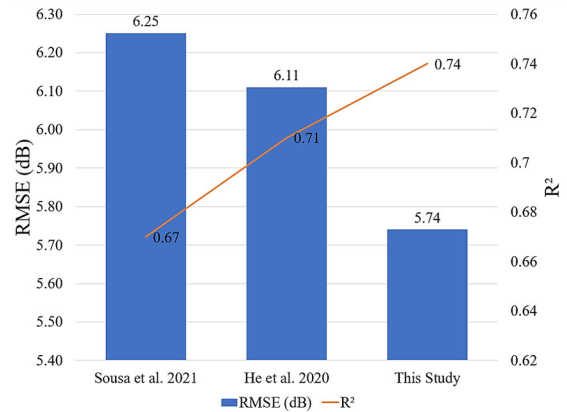**FIGURE 9.** The optimum hyperparameter settings for the RF model.

FIGURE 10. Comparison of RF model performance with the previous works.

massive dataset sizes, but it does not show a significant difference in prediction accuracy and model variability. In some cases, employing too many input parameters will make the ML learning algorithm more complex and deter the model's generalization capabilities. Besides, outliers' detection and removal activities must be executed properly before proceeding to the ML training process, especially when dealing with a massive number of datasets. Meanwhile, insufficient influential input parameters executed in[25] will limit the actual capabilities of the RF algorithm in producing optimal predictive performance. This is because, some other influential input parameters, such as signal propagation status LOS/NLOS are very significant in generating an accurate predictive result. Therefore, based on these comparison results, we can conclude that the five input parameters utilized in this study are adequate to produce an optimal RSRP prediction model.

VI. OPTIMIZATION

In this section, optimization approaches to existing 4G LTE networks using a developed prediction model are presented. Certain parameters such as height and tilt angle of the eNB antenna will be adjusted to forecast an improvement that can be achieved on poor RSRP readings. Poor RSRP readings from Cell ID 11, eNB 131163 will be utilized as a study
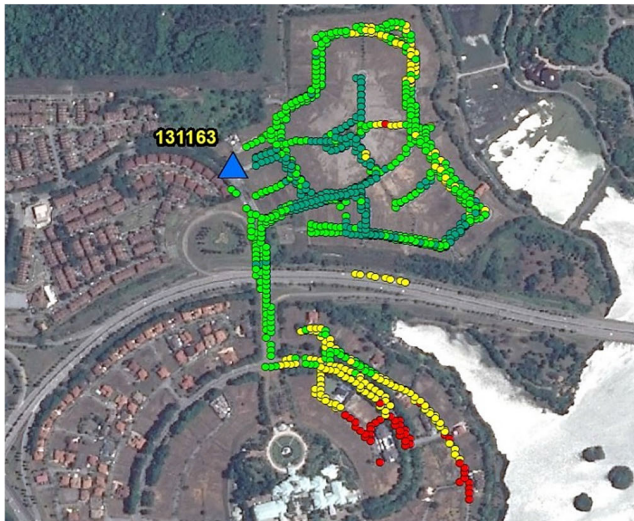
FIGURE 11. Original RSRP distribution for Cell ID 11, eNB 131163.

TABLE 7. Comparison of original and new setting.

Parameter	Original Setting	New Setting
Antenna Height	15 meter	30 meter
Antenna Tilt Angle	0°	3°

TABLE 8. RSRP performance before and after optimization.

RSRP Category	Before Optimization	After Optimization
Excellent	363	330
Good	633	671
Fair	219	253
Poor	45	6

FIGURE 12. RSRP distribution after optimization.

sample, as shown in Fig. 11. A total of 45 out of 1,260 point locations identified received poor RSRP readings.

Table 7 shows the original and new setting of Cell ID 11, eNB 131163. Over 100 prediction iterations were performed

per configuration parameter using the proposed model to obtain these new settings. Accordingly, different antenna height and tilt configuration values were utilized in each prediction iteration. The configuration values that resulted in the best coverage were then selected as optimal new settings. Model parameter inputs such as D_3, H_R and Tilt must be adjusted accordingly, while F_q and Obs are considered unchanged.

AS a result, an improvement in RSRP reading is shown in Table 8 and Fig. 12

It can be concluded that adjustment on both height and tilt angle of the antenna is able to improve the level of received signal strength at 39-point locations (86.7experienced poor receiving RSRP levels.

VII. CONCLUSION

This paper presented and examined several ML model' categories with various algorithms/kernels that aimed to predict mobile network performances through RSRP in multiband and multi-environment. For this purpose, ML models, including LR, ANN, SVM, RT, ET, and GPR, were applied and evaluated. The models were trained using measurement campaigns, carried out at 4G LTE frequency band in Malaysia, i.e., 1800 MHz and 2600 MHz, in diverse multi-environment around Putrajaya. The results showed that the GPR model is the most accurate model for RSRP prediction in terms of *RMSE* and R^2. However, due to its huge drawback on training times and prediction speed, the second-best model, which is the RF model, is highly recommended as the most practically employed ML-based model for developing a rigorous RSRP prognoses model in multiband and multi-environment. At the end of the article, optimization approaches on existing 4G LTE networks have been demonstrated by utilizing the capability of the developed ML-based RSRP prediction model. Finally, the future work is to train and test the RF algorithm with the real 5G NR measured data using the same approaches. Furthermore, the influence of spectrum bandwidth and UE position with respect to the front, side, and back lobe of antenna radiation patterns will be exploited. Besides, correlation with others network key performance parameters such as RSRQ will study and tested in future work.

REFERENCES

[1] K. Buchholz. (2021). *Where 5G Technology Has Been Deployed*. Statista Inc. Accessed: Nov. 25, 2021. [Online]. Available: https://www.statista.com/chart/23194/5g-networks-deployment-world-map/

[2] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, 4th Quart., 2018, doi: 10.1109/COMST.2018.2862141.

[3] S. Alraih, R. Nordin, I. Shayea, N. F. Abdullah, and A. Alhammadi, "Ping-pong handover effect reduction in 5G and beyond networks," in *Proc. IEEE Microw. Theory Techn. Wireless Commun. (MTTW)*, Oct. 2021, pp. 97–101, doi: 10.1109/MTTW53539.2021.9607205.

[4] *The Global Economic Impact of 5G*, PwC, London, U.K., Jul. 2020.

[5] I. Shayea, M. Ergen, M. H. Azmi, S. A. Colak, R. Nordin, and Y. I. Daradkeh, "Key challenges, drivers and solutions for mobility management in 5G networks: A survey," *IEEE Access*, vol. 8, pp. 172534–172552, 2020, doi: 10.1109/ACCESS.2020.3023802.

[6] N. F. Abdullah and M. L. H. Omar, "Millimeter wave performance evaluation in open areas and suburban Putrajaya," in *Proc. 4th Int. Conf. Electr., Electron. Syst. Eng. (ICEESE)*, Nov. 2018, pp. 1–5, doi: 10.1109/ICEESE.2018.8703562.

[7] C. A. Ionescu, M. T. Fülöp, D. I. Topor, S. Căpuşneanu, T. O. Breaz, S. G. Stănescu, and M. D. Coman, "The new era of business digitization through the implementation of 5G technology in Romania," *Sustainability*, vol. 13, no. 23, p. 13401, Dec. 2021, doi: 10.3390/su132313401.

[8] A. M. Ramly, N. F. Abdullah, and R. Nordin, "Cross-layer design and performance analysis for ultra-reliable factory of the future based on 5G mobile networks," *IEEE Access*, vol. 9, pp. 68161–68175, 2021, doi: 10.1109/ACCESS.2021.3078165.

[9] M. N. Mahdi, K. S. Mohamed, A. R. Ahmad, and M. A. Subhi, "The vision of 5G and cell-free communication networks in Malaysia," in *Proc. 8th Int. Conf. Inf. Technol. Multimedia (ICIMU)*, Aug. 2020, pp. 156–161, doi: 10.1109/ICIMU49871.2020.9243435.

[10] K. Buchholz. (2020). *5G Set to Beat Adoption Record for New Tech*. Statista Inc. Accessed: Nov. 26, 2021. [Online]. Available: https://www.statista.com/chart/22000/years-until-adoption-by-one-billion-users-selected-technologies/

[11] *5G Services Market Size, Share & Trends Analysis Report*, Grand View Research, Pune, India, 2021.

[12] H. Singh, S. Gupta, C. Dhawan, and A. Mishra, "Path loss prediction in smart campus environment: Machine learning-based approaches," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5, doi: 10.1109/VTC2020-Spring48590.2020.9129444.

[13] 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, doi: 10.1109/ACCESS.2020.3003247.

[14] A. B. M. D. S. Hossain. (Jul. 27, 2021). *Deployment Challenges of 5G Mobile Communication Infrastructure*. Daily Sun, Dhaka, Bangladesh. Accessed: Nov. 27, 2021. [Online]. Available: https://www.daily-sun.com/printversion/details/566669/Deployment-Challenges-of-5G-Mobile-Communication-Infrastructure

[15] A. Gupta, K. Ghanshala, and R. C. Joshi, "Machine learning classifier approach with Gaussian process, ensemble boosted trees, SVM, and linear regression for 5G signal coverage mapping," *Int. J. Interact. Multimedia Artif. Intell.*, vol. 6, no. 6, p. 156, 2021, doi: 10.9781/ijimai.2021.03.004.

[16] 3GPP. (2021). *Advanced Plans for 5G*. Accessed: Dec. 29, 2021. [Online]. Available: https://www.3gpp.org/news-events/2210-advanced_5g

[17] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60–69, Sep. 2019, doi: 10.1109/MVT.2019.2921627.

[18] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine learning-based methods for path loss prediction in urban environment for LTE networks," in *Proc. 16th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2020, pp. 1–6, doi: 10.1109/WiMob50308.2020.9253369.

[19] S. Ojo, A. Imoize, and D. Alienyi, "Radial basis function neural network path loss prediction model for LTE networks in multitransmitter signal propagation environments," *Int. J. Commun. Syst.*, vol. 34, no. 3, pp. 1–26, Feb. 2021, doi: 10.1002/dac.4680.

[20] S. Aldossari and K.-C. Chen, "Predicting the path loss of wireless channel models using machine learning techniques in mmWave urban communications," in *Proc. 22nd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Nov. 2019, pp. 1–6, doi: 10.1109/WPMC48795.2019.9096057.

[21] I. S. Popool, S. Misra, and A. A. Atayero, "Outdoor path loss predictions based on extreme learning machine," *Wireless Pers. Commun.*, vol. 99, no. 1, pp. 441–460, Mar. 2018, doi: 10.1007/s11277-017-5119-x.

[22] R. He, Y. Gong, W. Bai, Y. Li, and X. Wang, "Random forests based path loss prediction in mobile communication systems," in *Proc. IEEE 6th Int. Conf. Comput. Commun. (ICCC)*, Dec. 2020, pp. 1246–1250, doi: 10.1109/ICCC51575.2020.9344905.

[23] U. Masood, H. Farooq, and A. Imran, "A machine learning based 3D propagation model for intelligent future cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–5, doi: 10.1109/GLOBECOM38437.2019.9014187.

[24] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, "Neural networks and random forests: A comparison regarding prediction of propagation path loss for NB-IoT networks," in *Proc. 8th Int. Conf. Modern Circuits Syst. Technol. (MOCAST)*, May 2019, pp. 1–4, doi: 10.1109/MOCAST.2019.8741751.

[25] M. Sousa, A. Alves, P. Vieira, M. P. Queluz, and A. Rodrigues, "Analysis and optimization of 5G coverage predictions using a beam-forming antenna model and real drive test measurements," *IEEE Access*, vol. 9, pp. 101787–101808, 2021, doi: 10.1109/ACCESS.2021.3097633.

[26] N. Moraitis, L. Tsipi, D. Vouyioukas, A. Gkioni, and S. Louvros, "Performance evaluation of machine learning methods for path loss prediction in rural environment at 3.7 GHz," *Wireless Netw.*, vol. 27, no. 6, pp. 4169–4188, Aug. 2021, doi: 10.1007/s11276-021-02682-3.

[27] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path loss prediction based on machine learning: Principle, method, and data expansion," *Appl. Sci.*, vol. 9, p. 1908, May 2019, doi: 10.3390/app9091908.

[28] H.-S. Jo, C. Park, E. Lee, H. K. Choi, and J. Park, "Path loss prediction based on machine learning techniques: Principal component analysis, artificial neural network, and Gaussian process," *Sensors*, vol. 20, no. 7, p. 1927, Mar. 2020, doi: 10.3390/s20071927.

[29] B. Cha and S.-K. Noh, "Learning using LTE RSRP and NARNET in the same indoor area," in *Proc. 23rd Int. Comput. Sci. Eng. Conf. (ICSEC)*, Oct. 2019, pp. 261–264, doi: 10.1109/ICSEC47112.2019.8974774.

[30] T. Kim, K. Ko, I. Hwang, D. Hong, S. Choi, and H. Wang, "RSRP-based Doppler shift estimator using machine learning in high-speed train systems," *IEEE Trans. Veh. Technol.*, vol. 70, no. 1, pp. 371–380, Jan. 2021, doi: 10.1109/TVT.2020.3044175.

[31] M. F. A. Fauzi. (2021). *Machine Learning Training Dataset*. [Online]. Available: https://github.com/fazuwanfauzi/4G-LTE-RSRP-Prediction-Model/blob/main/Dataset.csv

[32] J. Wen, Y. Zhang, G. Yang, Z. He, and W. Zhang, "Path loss prediction based on machine learning methods for aircraft cabin environments," *IEEE Access*, vol. 7, pp. 159251–159261, 2019, doi: 10.1109/ACCESS.2019.2950634.

[33] R. Raj, A. Kulkarni, A. Seetharam, and A. Ramesh, "Wireless channel quality prediction using sparse Gaussian conditional random fields," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–6, doi: 10.1109/ccnc49032.2021.9369651.

[34] U. Lavanya, S. Mupparaju, P. Patnala, P. R. Anugu, and S. Surendran, "Model selection for path loss prediction in wireless networks," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Jul. 2020, pp. 1490–1493, doi: 10.1109/ICCSP48568.2020.9182186.

[35] S. Ray, "A quick review of machine learning algorithms," in *Proc. Int. Conf. Mach. Learn., Big Data, Cloud Parallel Comput. (COMITCon)*, Feb. 2019, pp. 35–39, doi: 10.1109/COMITCon.2019.8862451.

[36] D. F. S. Fernandes, A. Raimundo, F. Cercas, P. J. A. Sebastiao, R. Dinis, and L. S. Ferreira, "Comparison of artificial intelligence and semi-empirical methodologies for estimation of coverage in mobile networks," *IEEE Access*, vol. 8, pp. 139803–139812, 2020, doi: 10.1109/ACCESS.2020.3013036.

[37] S. I. Popoola, A. Jefia, A. A. Atayero, O. Kingsley, and N. Faruk, "Determination of neural network parameters for path loss prediction in very high frequency wireless channel," *IEEE Access*, vol. 7, pp. 150462–150483, 2019, doi: 10.1109/ACCESS.2019.2947009.

[38] E. Ostlin, H.-J. Zepernick, and H. Suzuki, "Macrocell path-loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, Jul. 2010, doi: 10.1109/TVT.2010.2050502.

[39] H. Elsherbiny, H. M. Abbas, H. Abou-zeid, H. S. Hassanein, and A. Noureldin, "4G LTE network throughput modelling and prediction," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–6, doi: 10.1109/GLOBECOM42002.2020.9322410.

[40] J. M. Roogi, "A quick review of ML algorithms," in *Proc. 6th Int. Conf. Electron. Commun. Electron. Syst. (ICCES)*, Jul. 2021, pp. 1–5, doi: 10.1109/ICCES51350.2021.9488982.

[41] MathWorks. (2021). *Train Regression Trees Using Regression Learner App*. MATLAB Help Center. Accessed: Nov. 26, 2021. [Online]. Available: https://www.mathworks.com/help/stats/train-regression-trees-using-regression-learner-app.html

[42] S. Duangsuwan and M. M. Maw, "Comparison of path loss prediction models for UAV and IoT air-to-ground communication system in rural precision farming environment," *J. Commun.*, vol. 16, no. 2, pp. 60–66, 2021, doi: 10.12720/jcm.16.2.60-66.

[43] A. Narayanan, E. Ramadan, R. Mehta, X. Hu, Q. Liu, R. A. K. Fezeu, U. K. Dayalan, S. Verma, P. Ji, T. Li, F. Qian, and Z.-L. Zhang, "Lumos5G: Mapping and predicting commercial mmWave 5G throughput," in *Proc. ACM Internet Meas. Conf.*, Oct. 2020, pp. 176–193, doi: 10.1145/3419394.3423629.

[44] MathWorks. (2021). *Regression Tree Ensembles*. MATLAB Help Center. Accessed: Nov. 26, 2021. [Online]. Available: https://www.mathworks.com/help/stats/regression-tree-ensembles.html

[45] D. Minovski, C. Ahlund, K. Mitra, and P. Johansson, "Analysis and estimation of video QoE in wireless cellular networks using machine learning," in *Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX)*, Jun. 2019, pp. 1–6, doi: 10.1109/QoMEX.2019.8743281.

[46] M. S. H. Lipu, A. Ayob, M. H. M. Saad, A. Hussain, M. A. Hannan, and M. Faisal, "State of charge estimation for lithium-ion battery based on random forests technique with gravitational search algorithm," in *Proc. IEEE PES Asia–Pacific Power Energy Eng. Conf. (APPEEC)*, Oct. 2018, pp. 45–50, doi: 10.1109/APPEEC.2018.8566648.

[47] N. H. Anh, N. K. Hung, and N. Van Khang, "Received signal strength prediction using Gaussian process," in *Proc. Int. Conf. Recent Adv. Signal Process., Telecommun. Comput. (SigTelCom)*, Jan. 2017, pp. 94–97, doi: 10.1109/SIGTELCOM.2017.7849802.

[48] N. Zhang, J. Xiong, J. Zhong, and K. Leatham, "Gaussian process regression method for classification for high-dimensional data with limited samples," in *Proc. 8th Int. Conf. Inf. Sci. Technol. (ICIST)*, 2018, pp. 358–363, doi: 10.1109/ICIST.2018.8426077.

[49] S. Hong and Z. Zhou, "Application of Gaussian process regression for bearing degradation assessment," in *Proc. 6th Int. Conf. New Trends Inf. Sci., Service Sci. Data Mining (ISSDM)*, Oct. 2012, pp. 644–648.

[50] S. G. Watza and E. W. Frew, "Online hybrid RF propagation model for communication-aware sUAS relay application," in *Proc. AIAA Inf. Syst.-AIAA Infotech@ Aerosp.*, Jan. 2018, Art. no. 209989, doi: 10.2514/6.2018-2141.

[51] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. 33rd Int. Conf. Mach. Learn. (ICML)*, vol. 3, 2016, pp. 1651–1660.

[52] H. A. H. Alobaidy, J. S. Mandeep, R. Nordin, N. F. Abdullah, C. G. Wei, and M. L. S. Soon, "Real-world evaluation of power consumption and performance of NB-IoT in Malaysia," *IEEE Internet Things J.*, early access, Nov. 29, 2021, doi: 10.1109/JIOT.2021.3131160.

[53] Malaysian Communications and Multimedia Commission (MCMC). (2018). *5G Test Bed: Call For Collaboration*. Accessed: Nov. 19, 2021. [Online]. Available: https://www.mcmc.gov.my/en/media/announcements/5g-test-bed-call-for-collaboration

[54] Malaysian Communications and Multimedia Commission (MCMC). *5G Malaysia Demonstration Project?: FAQs*. Accessed: Nov. 19, 2021. [Online]. Available: https://www.nfcp.my/Nfcp/media/Docs/FAQs_5G-Demonstration-Projects.pdf

[55] Gyokov Solutions. *G-NetTrack Pro*. Accessed: Nov. 29, 2021. [Online]. Available: https://gyokovsolutions.com/

[56] Kliksob Network. *4G LTE Switcher*. Accessed: Nov. 29, 2021. [Online]. Available: https://play.google.com/store/apps/details?id=com.kliksob.forgswitch&hl=en&gl=U.S

[57] V. Raida, P. Svoboda, and M. Rupp, "On the inappropriateness of static measurements for benchmarking in wireless networks," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–5, doi: 10.1109/VTC2020-Spring48590.2020.9128867.

[58] *Online RF Planning Software—Cloud-RF*. [Online]. Available: https://cloudrf.com/

[59] *OpenStreetMap*. Accessed: Nov. 29, 2021. [Online]. Available: https://www.openstreetmap.org/#map=6/4.116/109.455

[60] *5G; Study on Channel Model for Frequencies From 0.5 to 100 GHz*, 3GPP document TR 38.901, version 14.0.0 Release 14, 2017.

MOHD FAZUWAN AHMAD FAUZI received the B.Eng. degree in electronics from Multimedia University (MMU), in 2005. He is currently pursuing the M.Sc. degree with the Department of Electrical, Electronics and Systems Engineering, Universiti Kebangsaan Malaysia. Since 2007, he has been working as a Researcher at the Malaysian Space Agency (MYSA), a government research institute. His research interests include wireless communications, data science, and applied machine learning algorithms for mobile networks.

ROSDIADEE NORDIN (Senior Member, IEEE) received the B.Eng. degree from the Universiti Kebangsaan Malaysia, in 2001, and the Ph.D. degree from the University of Bristol, U.K., in 2011. He is currently a Professor with the Department of Electrical, Electronic and Systems Engineering, Universiti Kebangsaan Malaysia, teaching different subjects related to wireless networks and mobile communications.

NOR FADZILAH ABDULLAH (Member, IEEE) received the B.Sc. degree in electrical and electronics from the Universiti Teknologi Malaysia, in 2001, the M.Sc. degree (Hons.) in communications engineering from The University of Manchester, U.K., in 2003, and the Ph.D. degree in electrical and electronic engineering from the University of Bristol, U.K., in 2012. She is currently an Associate Professor with the Universiti Kebangsaan Malaysia, Selangor, Malaysia. Her research interests include 5G, millimeter wave, LTE-A, vehicular networks, massive MIMO, space time coding, fountain codes, and channel propagation modeling and estimation.

HAIDER A. H. ALOBAIDY (Graduate Student Member, IEEE) received the M.Sc. degree in electrical engineering/electronics and communication from the Faculty of Engineering, Al-Mustansiriyah University, Iraq, in 2016. He is currently pursuing the Ph.D. degree with the Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM). His research interests include wireless sensor networks, wireless communications, the IoT, and channel propagation modeling and estimation.

• • •