IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

IEEE SYSTEMS, MAN AND CYBERNETICS SOCIETY SECTION

Received 25 November 2022, accepted 30 December 2022, date of publication 5 January 2023, date of current version 11 January 2023.

*Digital Object Identifier 10.1109/ACCESS.2023.3234566*

APPLIED RESEARCH

# Machine Learning-Based Online Coverage Estimator (MLOE): Advancing Mobile Network Planning and Optimization

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

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

Corresponding author: Nor Fadzilah Abdullah (fadzilah.abdullah@ukm.edu.my)

This work was supported in part by ISIF Asia Foundation under grant code M-202106-00113 (UKM reference: KK-2021-020).

**ABSTRACT** Nowadays, the dependency on high-performance digital mobile connectivity is not limited to human usage but also the intelligent objects increasingly deployed to serve the needs of Internet of Things (IoT) applications. However, the current network planning technique limitation has constrained the real potential of mobile digital connectivity development. This situation has hindered sustainable Internet-oriented economic and technological development. The 3rd generation partnership project (3GPP), through its specification release 18 (Rel.18), has included and leveraged the potential capabilities of machine learning (ML) technologies in advanced mobile network planning. The main objective is to enhance mobile network planning performance and reduce complexity. To materialize this aim, we propose a novel ML-based Online coverage Estimator (MLOE) tool developed based on Random Forest (RF) ML algorithm. It uses seven unique features to predict the mobile network performance through reference signal received power (RSRP). Accordingly, the results showed that MLOE outperformed traditional empirical techniques and previous works. The final trained RF algorithm has achieved an outstanding root mean square error (*RMSE*) of 2.65 dB and a coefficient of determination ($R^2$) of 0.93. With the dynamic and fast-growing mobile technology, MLOE has been deployed on an online platform using MATLAB® Web App Server, which offers a modular and scalable architecture.

**INDEX TERMS** Machine learning, MATLAB, mobile networks, path loss, received signal strength indicator, RSRP, web application.

## I. INTRODUCTION

Despite the recent technological advances, issues related to unsatisfactory mobile network services are still challenging, as highlighted by the Malaysian Communications and Multimedia Commission (MCMC) [1]. Among the contributing factors to the latter issue is the current network planning techniques' limitation. According to [2], [3], [4], and [5], the traditional empirical techniques primarily applied in the industry are inaccurate. Meanwhile, the deterministic techniques are not practical to apply on real-world operational scales due to their complexity, requirement for high-resolution topographic maps, intense reference information, and high computing power.

Aiming to address the latter issues, 3GPP, through its Rel.18, has included and leveraged the potential capabilities of ML techniques to enhance mobile network planning performance and reduce complexity [6]. In this regard, four main objectives were outlined. The first is focused on identifying a common ML framework, including the functional

The associate editor coordinating the review of this manuscript and approving it for publication was Huiyan Zhang.

requirements of ML architecture. Secondly, to identify areas where ML could improve the performance of mobile network planning functions. Thirdly, to identify what is required for an adequate ML model characterization and description, establishing proper notation. Then, evaluate ML-based techniques to understand the attainable gains and complexity requirements.

In principle, ML techniques are less complex than traditional empirical methods in producing a propagation model. In the former method, the researcher has to find the appropriate rules and algorithms to produce the required forecast output quality. Because of this, most propagation models based on traditional empirical techniques are inflexible due to the constraints faced [5]. On the other hand, in the ML technique, this process is automatically done by the computer system based on the characteristics of input parameters and response variables that have been set. Fig. 1 shows the comparison of the working principles of ML techniques and traditional empirical techniques in producing a prediction model.

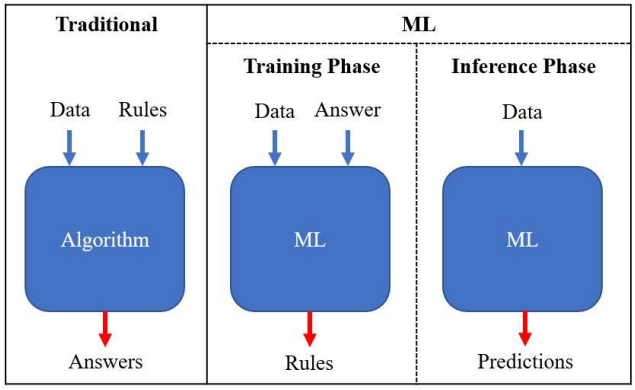


**FIGURE 1.** Traditional algorithms versus ML Models.

Accordingly, this study will review and identify the most practical and optimum ML algorithm for developing MLOE for RSRP-based prediction. RSRP is one of the 4G and 5G networks' key performance indicators (KPI) that telco companies utilize to understand and evaluate the performance and coverage of their network [7], [8]. While the specification of RSRP in 4G and 5G are not similar, RSRP, in both technologies, serves the same function of performing cell selection/reselection and handover process [9].

In the meantime, this study will also examine several features to evaluate their suitability for RSRP prediction. Finally, we have utilized the MATLAB Web App server for the MLOE deployment, which is publicly accessible via the Internet. Fig. 2 describes the overall concept of execution of this study.

As Malaysia's 5G mobile network is still under development and testing, this study focused on utilizing the existing 4G network. This is also in line with the direction of Malaysia's government through the *Jalinan Digital Negara* (JENDELA) plan. The plan has enforced the 3G network services termination by late 2021 to empower 4G networks

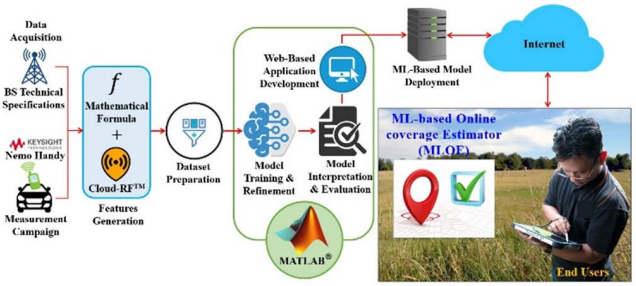


**FIGURE 2.** Overall execution concept of this study.

for bridging the digital divide, especially between rural and urban areas [10].

As such, the contributions of this study can be summarized as follows

- Describes an extensive measurement campaign consisting of a physical layer dataset with over 14,464 data samples collected in various outdoor environments using Keysight Nemo Handy drive test solution, available publicly through an open-source GitHub repository at [11].
- Insight into the trained Random Forest (RF) model behaviour is gained through model-agnostic methods, especially to validate the impact of utilizing new types of features, i.e., $AZ_{offset}$ and $Tilt_{offset}$ for RSRP prediction.
- The performance of the final trained RF model is evaluated against traditional empirical techniques and previous works done by [5], [12], [13], [14], and [15].
- Describes a comprehensive MLOE development methodology for mobile network planning and optimization where the final proposed tool can be accessed online freely at [16].

The rest of the paper is organized as follows: Section II discusses the recent progress in ML-based propagation modelling. Section III presents the methodology of the study.

The results of the study are discussed in Section IV. Finally, Section V concludes the article.

## II. RECENT PROGRESS IN ML-BASED PROPAGATION MODELLING

Recently, ML techniques have been actively explored by researchers from various fields, including mobile telecommunications [17], [18], [19], [20]. The latter is because ML techniques that focus on data and algorithms have proven their capabilities in many fields of application, such as medicine, automotive, economics, banking and many other fields [21], [22], [23], [24]. In addition, ML-based models can improve their accuracy over time without having to be specifically programmed [25]. Therefore, it coincides with the development trend of mobile network technology which is dynamic and rapidly evolving [26], [27], [28], [29].

Predicting mobile network coverage based on specific features is categorized as a regression-type supervised machine learning technique [12], [30]. This field of study, especially for the ground-to-ground (G2G) mobile networks, has been

explored by [5], [12], [13], [14], and [15] using a real-world dataset. In the latter studies, most ML models' prediction accuracy was around 4 dB to 8 dB, depending on the type of environment studied. While each study reported different findings, it was found that the Random Forest (RF) ML algorithm consistently showed the best prediction performance. RF was also found to be more efficient regarding training duration and prediction speed than other ML models such as Artificial Neural Network (ANN), Support Vector Machine (SVM) and Gaussian process regression (GPR) [12], [15]. On the other hand, RF is also very well known for its robustness and powerful capabilities, especially not susceptible to sample and feature disturbances also noise [12], [13], [15]. Hence, RF has been selected as the basic framework in this study which is in line with the study objectives to develop a practical and optimal RSRP prediction model for use in real-world operations.

In [15], we have listed the features utilized in the previous works. According to domain knowledge, among identified features that highly affect the performance of mobile network signal propagation are the antenna tilt angle, azimuth angle, position and location [31]. Therefore, in Table 1, we have assessed and summarized the pros and cons of the features utilized in previous works, which finally led to determining the features to be utilized in this study.

Previous works never tested the trained RF model outside the study area. This may be because the ML technique, which falls under the empirical category, tends to be inaccurate when applied outside the study area. Therefore, this study will test the performance of the final trained RF model inside and outside the study area and compare it with the prediction results from the traditional empirical propagation models.

Most of the time, the wireless planning tools in the desktop version are inflexible. This is because it was developed based on a conventional framework. However, the innovation of mobile network planning through an online platform, such as pioneered by CloudRF [32], is a game changer in telecommunications. In addition to its advantage of being accessible anywhere, the flexible and scalable development architecture makes upgrading the system much easier and simpler. At the same time, sharing the latest update or information with end users will be more effective.

Meanwhile, the burden of data processing and analysis is entirely on the server side, which gives the user a huge advantage. End users do not need to worry about processing power and storage capacity requirements. They only need to ensure the Internet connectivity is stable. This kind of development framework is more user-oriented. Therefore, the service package offered is usually suited to most end users' preferences and financial capabilities. The identical framework has been applied by Internet giant companies such as Amazon, Microsoft, and Google in their cloud computing services [33], [34], [35].

However, a survey of research articles related to ML found that less than 10% of studies discuss the deployment of ML models, and far fewer of them successfully deploy ML models into real-world applications [36]. This is because deploying ML models in a real operational environment is critical and challenging [37]. ML models successfully deployed at the production level can only increase operational efficiency or develop new value propositions [38].

**TABLE 1. A summary of the pros and cons of features utilized in previous works.**

| Ref. | Features | Review/Comment |
|---|---|---|
| [13], [14] | • Distance transmitter ($T_x$) to the receiver ($R_x$)<br>• Frequency | Based on domain knowledge, the separation distance between $\boldsymbol{T_x}$ and $\boldsymbol{R_x}$ also frequency band is considered an important feature in predicting the coverage of mobile networks. However, without other important information, such as $\boldsymbol{T_x}$ antenna azimuth and tilt angle, the position of $\boldsymbol{R_x}$ relatively to the $\boldsymbol{T_x}$ at various planes will deter the overall model prediction performance. This is because most 4G and 5G $\boldsymbol{T_x}$ antennas widely deployed are directional types. |
| [12] | • Distance $T_x$ to $R_x$<br>• Height of $T_x$ from above sea level (ASL)<br>• Height of $R_x$ (ASL)<br>• LOS/NLOS | This study utilised the height of $\boldsymbol{T_x}$ and $\boldsymbol{R_x}$ as features which are in line with domain knowledge. Meanwhile, using LOS/NLOS status is also a good approach. This is because the signal propagation status highly influences the mobile signal performance. Like the abovementioned works, without an antenna azimuth and tilt angle information, the overall ML-based model prediction performance will be deterred because most of the 4G and 5G $\boldsymbol{T_x}$ antennas are directional types. |
| [5] | • Distance $T_x$ to $R_x$<br>• Frequency<br>• Height difference $T_x$ to $R_x$<br>• $T_x$ tilt angle<br>• $T_x$ azimuth angle<br>• Transmit power<br>• Clutter & building heights<br>• Vertical distance of $R_x$ from the boresight of the $T_x$ antenna | The features used in this study are among the most comprehensive compared to the three works above. However, lack of information related to the horizontal distance of $\boldsymbol{R_x}$ from the boresight of the $\boldsymbol{T_x}$ antenna seen distorted the true potential of this model. In the meantime, this study does not address the other influenced features related to the environment's characteristics, such as undulating ground surface. |
| [15] | • Distance $T_x$ to $R_x$<br>• Frequency<br>• Height ratio $T_x$ to $R_x$ (ASL)<br>• The $R_x$ elevation angle<br>• LOS/NLOS | This is our previous work. With limited information, we tried to maintain the right features utilised in other previous works. In this work, we introduced a new feature which is the height ratio of $\boldsymbol{T_x}$ to $\boldsymbol{R_x}$ at ASL. We found that the use of this feature should be avoided as it leads to inconclusive results in site-specific modelling. |

## III. METHODOLOGY

To develop an accurate RSRP prediction model, the features must be able to describe precisely the receiver ($\boldsymbol{R_x}$) location in reference to the location of the transmitter ($\boldsymbol{T_x}$) antenna. Other than that, the features also must be able to describe the signal propagation status and the characteristics of the operational environment. Therefore, the level of signal attenuation experienced before arriving at the $\boldsymbol{R_x}$ location can be

predicted more precisely. In this study, the MLOE development activities were carried out based on the workflow described in Fig. 3. Explanation regarding each step in the following workflow is briefly described as follows:

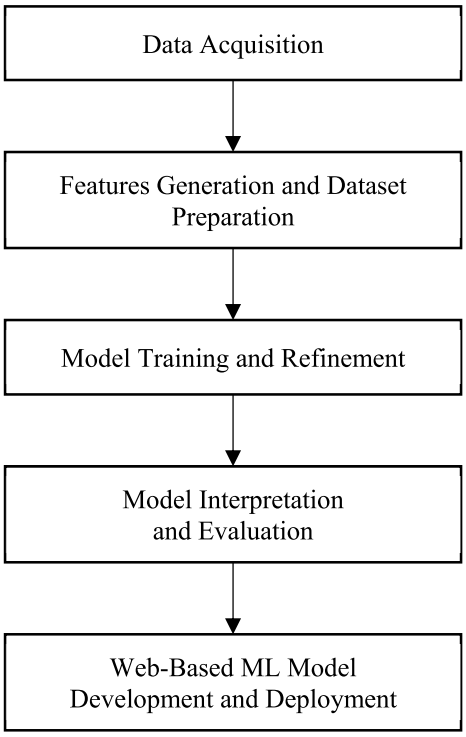


**FIGURE 3.** MLOE development workflow.

### A. DATA ACQUISITION

Two types of raw data were acquired, which are: (i) the 4G mobile network base station (BS) technical specifications; and (ii) the real-world 4G RSRP data. The real-world 4G RSRP data was obtained through measurement campaign activities conducted in 14 areas around the Klang Valley, Malaysia, representing dense urban, urban, suburban, and open area environments using hardware and software described in Table 2. The measurement campaign was conducted at a vehicle speed below 40 km/h to minimize the fast-fading effect due to the Doppler shifts [4], while maintaining the speed limit regulation in respective areas.

The measurement campaign was done for two purposes: (i) to generate the RF model training dataset and (ii) to generate a test dataset for the final trained RF model. For model training, the data collected in Putrajaya are only used (refer to Fig. 4). It consists of 10 transceiver BS antennas. Putrajaya is a unique region with multiple environmental categories [39], [40]. Meanwhile, for final trained RF model testing and evaluation purpose, 12 transceiver BS located inside and outside of Putrajaya have been identified (refer to Table 3 ). The details information regarding the test location and distribution of test points are available in the open-source GitHub repository at [41].

**TABLE 2.** Parameters of measurement campaign setup.

| | |
|---|---|
| **HARDWARE:** | |
| Smartphone | Samsung S21 Ultra 5G |
| User Equipment (UE) | Category 20 |
| 3GPP Release Standards | Release 16 |
| **SOFTWARE:** | |
| Nemo Handy [42] | Keysight android-based solution that enables measuring wireless diagnostics information of air interface and mobile application quality-of-service (QoS) and quality-of-experience (QoE) |

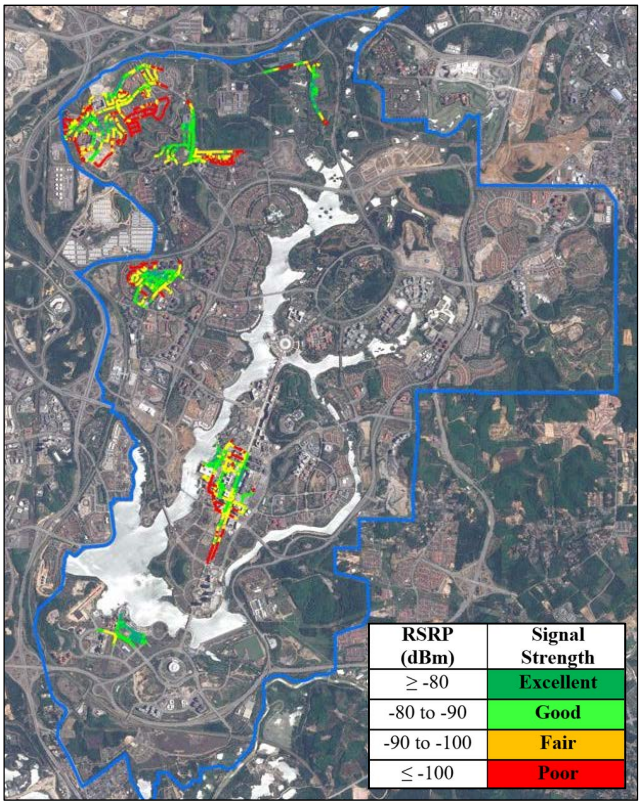


**FIGURE 4.** Distribution of drive test data around Putrajaya.

Although there are several test locations outside Putrajaya with different types of land use and environmental characteristics, the selection of these locations is purposely to analyse the trained RF model generalization capability. This model prediction performance would then be compared to the traditional empirical propagation models. These tests would also identify whether the final trained RF model experiences overfitting, where the trained model tends to produce very poor prediction performance on the test area located outside the study area. Thus, the assumption is that if the prediction performance inside and outside training areas is minimal, then the trained model would have the potential to provide reliable prediction capabilities in Malaysia's harsh tropical/irregular-terrain environments and other similar context areas.

**TABLE 3.** List of final trained RF model test locations.

| No. | Test Location | Area Profile |
|---|---|---|
| **Located inside the study area:** | | |
| 1. | Presint 3, Putrajaya | Urban |
| 2. | Presint 11, Putrajaya | Suburban |
| 3. | Presint 5, Putrajaya | Open Area |
| **Located outside the study area:** | | |
| 4. | Bukit Bintang, Kuala Lumpur | Dense Urban |
| 5. | SOGO, Kuala Lumpur | Dense Urban |
| 6. | Cyberjaya, Selangor | Urban |
| 7. | Bandar Seri Putra, Selangor | Suburban |
| 8. | Sungai Ramal Dalam, Selangor | Suburban |
| 9. | Putra Perdana, Selangor | Suburban |
| 10. | KLIA Mosque, Selangor | Open Area |
| 11. | Bandar Baru Bangi, Selangor | Open Area |
| 12. | Bandar Baru Nilai, Negeri Sembilan | Open Area |

### B. FEATURES GENERATION AND DATASET PREPARATION

To create the dataset for model training purposes, the raw data mentioned earlier was utilized to generate seven unique features as follows:

#### 1) DISTANCE

The distance ($\boldsymbol{D}$) is the separation distance in meters between BS and user equipment (UE) at the Euclidean plane. It is used to estimate the UE position referenced to the position of the BS antenna on the x-axis. The calculation of $\boldsymbol{D}$ is based on the spherical law of cosines [43], [44].

#### 2) AZIMUTH OFFSET ANGLE

The azimuth offset angle ($\boldsymbol{AZ_{offset}}$) represents the absolute angle formed at the horizontal plane between the BS antenna pointing direction and the location of UE. It is used to estimate the UE position referenced to the position of the BS antenna on the y-axis. In the meantime, $\boldsymbol{AZ_{offset}}$ is also used to estimate the UE position within the coverage area of the BS antenna main lobe in the horizontal plane. The received signal strength will be stronger if the UE position is near the BS antenna boresight [45], [46]. The calculation of $\boldsymbol{AZ_{offset}}$ is given in (1).

$$\boldsymbol{AZ_{offset} = |AZ_{ant} - AZ_{ue}|} \tag{1}$$

where, $\boldsymbol{AZ_{ant}}$ refers to the BS antenna azimuth angle. Meanwhile, $\boldsymbol{AZ_{ue}}$ refers to the UE azimuth angle calculated using the arctan2 function [47].

#### 3) ELEVATION ANGLE

The elevation angle ($\boldsymbol{ELEV}$) is referred to the angle formed between the horizontal plane and the observed line from UE to the BS antenna. $\boldsymbol{ELEV}$ is used to estimate the UE position with reference to the position of the BS antenna on the z-axis. Other than that, $\boldsymbol{ELEV}$ is also used to describe the characteristics of the UE's position on the undulating Earth's surface. Even at the same separation distance between the BS and the UE, the $\boldsymbol{ELEV}$ value will differ according to the actual position of the UE either in the highlands or in the valleys. The calculation of $\boldsymbol{ELEV}$ is given in (2).

$$\boldsymbol{ELEV = tan^{-1}\left((h_{ant} - h_{ue})/D\right)} \tag{2}$$

whereas $\boldsymbol{h_{ant}}$ is the antenna height and $\boldsymbol{h_{ue}}$ is the UE height. Both height values are referred to above sea level (ASL). $\boldsymbol{D}$ is the value of the separation distance between $\boldsymbol{T_x}$ and $\boldsymbol{R_x}$ as described in (1). The web-based radio planning tool called CloudRF [32], which is equipped with 10 meters resolution of terrain and clutter information, has been utilised to generate the values of $\boldsymbol{h_{ant}}$ and $\boldsymbol{h_{ue}}$.

#### 4) TILT OFFSET ANGLE

The tilt offset angle ($\boldsymbol{TILT_{offset}}$) is the angle formed between the boresight of the BS antenna and the observed line from UE to the BS antenna on the vertical plane. $\boldsymbol{TILT_{offset}}$ is used to estimate the UE position in the beam of the main lobe of the BS antenna on the vertical plane. Based on the Pythagorean theorem, the value of the observation angle from the BS antenna to the UE from the horizon is equal to $\boldsymbol{ELEV}$. Meanwhile $\boldsymbol{TILT_{ant}}$ is equal to the sum of the mechanical and electrical tilt angle of the BS antenna. Therefore, the calculation of $\boldsymbol{TILT_{offset}}$ is given in (3).

$$\boldsymbol{TILT_{offset} = |ELEV - TILT_{ant}|)} \tag{3}$$

#### 5) ENVIRONMENT CATEGORY (*CLASS*)

At the same specification of BS, the extent of mobile network coverage will vary according to its operating environment [48]. The transmitted signal will be reflected more frequently in urban areas compared to suburban and open areas. As a result, the signal attenuates much faster in urban areas. Because of that, the dataset in this study was classed accordingly into three categories: urban, suburban, and open areas. Even though the RF algorithm is insensitive to the dataset's variance [49], for computational simplicity, this category was coded as 1, 2 and 3 to represent urban, suburban, and open areas, respectively.

#### 6) FREQUENCY BANDS (*FQ*)

The frequency band, also called radio frequency, is the air interface medium that carries information from the transmitter (BS antenna) to the receiver (UE). Different frequency bands have different characteristics and capabilities. Low-frequency bands give larger coverage, but the transmitted signal capacity is low. In contrast, the high-frequency band can carry high-capacity applications such as video calls, online movies and more. But the downside is the limited coverage. In this study, two 4G mobile networks frequency bands are studied, i.e., 1800 MHz dan 2600 MHz.

#### 7) SIGNAL PROPAGATION STATUS (*OBS*)

The propagation of mobile signals is greatly affected by the surrounding objects. This is due to the nature of radio waves that will be scattered, reflected, refracted, or absorbed when interacting with any object in its path [50]. As a result, it affects the signal power level at the UE location. Therefore,

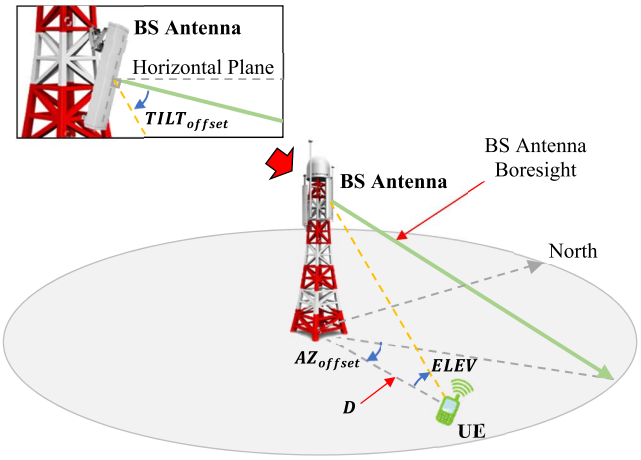


**FIGURE 5.** Position of calculated features.

the features related to signal propagation status were included in this study. Due to the limited information, the signal propagation status in this study was categorised roughly into two classes, which are line-of-sight (LOS) and non-line-of-sight (NLOS). This feature was extracted using the CloudRF radio planning tool, which is also embedded with high-resolution 3D building maps from OpenStreetMap [51]. The signals propagation status was coded as 0 and 1 to represent LOS and NLOS, respectively.

The abovementioned features are illustrated in Fig. 5 and summarized in Table 4 for a clear explanation and comparison.

Before proceeding to the model training and refinement activities, the training dataset must undergo a data-cleaning process to eliminate outliers. This cleaning process was carried out using an interquartile range (IQR) score [52]. A total of 16,310 points of the training dataset were used in this process. As a result, 1,846 outliers' data were detected and eliminated. The same IQR score is then utilised to clean the test dataset. The aim is to ensure that the final trained RF model is tested within the studied data range. Details about IQR scores used in this study are available publicly through an open-source GitHub repository [53].

### C. MODEL TRAINING AND REFINEMENT

MATLAB 2020a Regression Learner was used to train the RF model. The training process was performed using 10-fold cross-validation (CV), a resampling method for the training and validating process, mainly to prevent overfitting [5], [14], [15]. Evaluation matrix, Root-mean-square error (***RMSE***) and coefficient of determination ($\boldsymbol{R^2}$) were used to evaluate the performance of the trained model. The calculation of ***RMSE*** and $\boldsymbol{R^2}$ is given in (4) and (5), respectively.

$$\boldsymbol{RMSE} = \sqrt{\frac{1}{n_{sample}} \sum_{i=0}^{n_{sample}-1} \left|y_i - \hat{y}_i\right|^2} \tag{4}$$

**TABLE 4.** List of features and explanations.

| Features | Definition | Purpose |
|---|---|---|
| $D$ | The separation distance between BS and UE on the Euclidean plane | To estimate the position of UE away from the BS antenna on the x-axis |
| $AZ_{offset}$ | The absolute value of the angle formed between the BS antenna pointing direction and the location of UE on the horizontal plane | To estimate the UE position with reference to the position of the BS antenna on the y-axis. Also, to estimate the UE position in the main lobe of the BS antenna on the horizontal plane. |
| $ELEV$ | The angle formed between the horizontal plane and the observed line from UE to the BS antenna | To estimate the UE position with reference to the position of the BS antenna on the z-axis. It is also used to describe the characteristics of the UE's position on the undulating Earth's surface. |
| $TILT_{offset}$ | The absolute value of the angle formed between the BS antenna boresight and the observed line from UE on the BS antenna | To estimate the UE position in the beam of the main lobe of the BS antenna on the vertical plane. |
| $CLASS$ | The characteristics of mobile networks' operational environment, whether urban, suburban or open areas | To estimate the characteristics of the propagation channel between BS and UE |
| $FQ$ | The 4G spectrum bands in Malaysia as allocated by MCMC. 1800 MHz or 2600 MHz | To estimate the level of attenuation experienced by the transmitted signal from the BS to the UE according to different spectrum bands being utilized. |
| $OBS$ | The status of the signal propagation from the BS antenna to the UE, whether LOS or NLOS | To estimate the level of attenuation experienced by the BS transmitted signal before arriving at the UE due to the presence of obstacles |

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{i} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{i} \left(y_i - \bar{y}_i\right)^2}, \textbf{where } \bar{y} = \frac{1}{i} \sum_{i=1}^{i} y_i \tag{5}$$

where $n_{sample}$ is the total number of samples, $y_i$ is actual value, and $\hat{y}_i$ is the predictive value. According to [12], if the *RMSE* less than 7 dB, it is considered acceptable for urban environment. While 10 to 15 dB is acceptable for suburban and rural area.

During the model training session, important feature analysis is performed using the feature selection function available in MATLAB Regression Learner. This analysis is to identify and remove non-influential features. This kind of analysis also has been utilised by [5], [12], and [15] in their works.

Hyperparameter tuning is the final process in the model training session, where a model will be tuned in more detail to obtain the most optimal performance results. For this purpose,

the RF model was tuned using the Optimizable Ensemble function, also available in MATLAB Regression Learner and the results are shown in Fig. 6. The newly generated values of ***RMSE*** and $\boldsymbol{R^2}$ are equal to 2.65 dB and 0.93, respectively. Therefore, the accuracy of the final trained RF model was increased by 0.58 dB and model variability increased by 0.03. As a result, the final trained RF model hyperparameter settings have 499 learners, two minimum leaf sizes, and five predictors to sample.

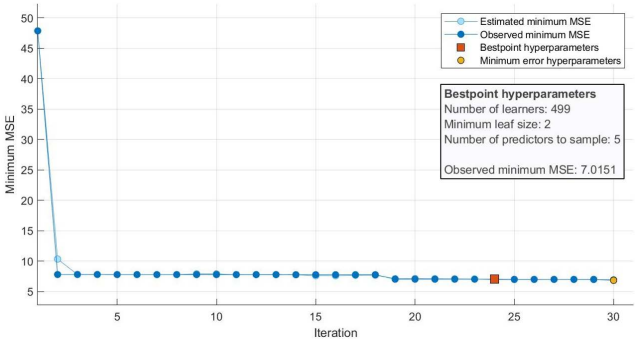


**FIGURE 6.** The final trained RF model hyperparameter settings.

The final trained RF model is then exported to the MATLAB workspace for interpretation and evaluation. In summary, the ML model training and refinement activities were implemented based on the flow chart shown in Fig. 7.

### D. MODEL INTERPRETATION AND EVALUATION

ML model like RF is often referred to as black box model because it is difficult to interpret how the model makes predictions [54], [55]. Interpretability tools help to overcome this issue and reveal how the features contribute to the predictions. In the meantime, interpretability tools can validate whether the model uses the correct principle for its predictions based on domain knowledge. Besides, it can also find model biases that are not immediately apparent.

In this study, we utilised a partial dependent plot (PDP), a type of Model-Agnostic Method, which aims to show the marginal effect that features have on the predicted outcome [54], [56]. PDP is a global interpretation tool that can explain how a trained model makes predictions for the entire data set. The partial dependence function for regression is defined as (6) [54]:

$$\hat{f}_{x_s}(x_s) = E_{x_c}\left[\hat{f}(x_s, x_c)\right] = \int \hat{f}(x_s, x_c)\, dP(x_c) \quad (6)$$

whereas $x_s$ are the features for which the partial dependence function should be plotted and $x_c$ are the other features used in the ML model $\hat{f}$.

From the PDP graph, we extract the X and Y values and calculate the standard deviation (SD) as defined in (7) [57]. The most important features will show higher SD values.

$$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} |A_i - \mu|^2} \quad (7)$$

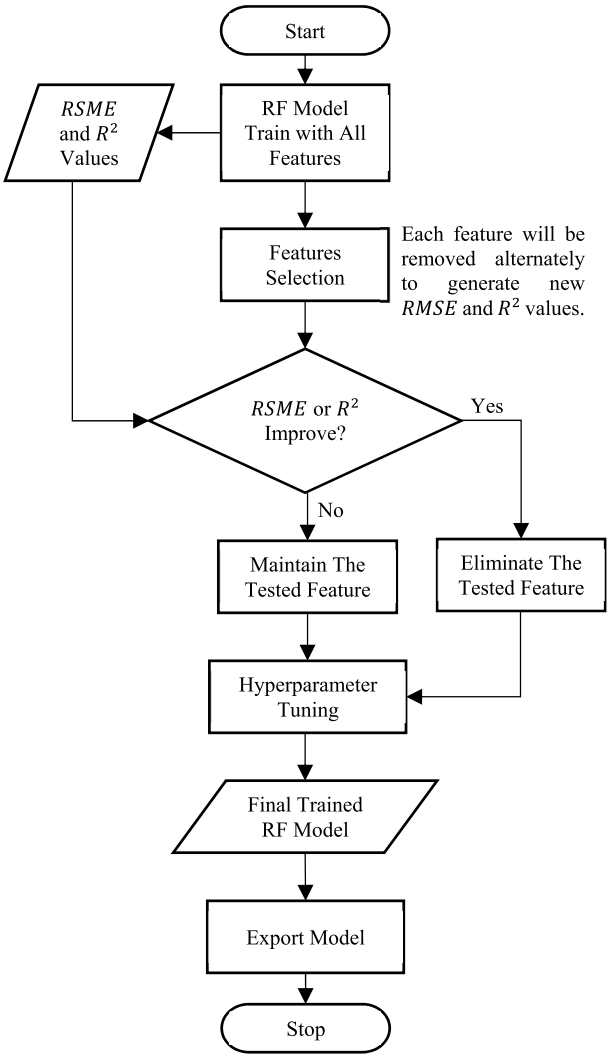


**FIGURE 7.** The flowchart of the implementation of model training and refinement activities conducted in this study.

where $\mu$ is the mean of $A$:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} A_i \quad (8)$$

Two approaches have been taken to assess and evaluate the performance of the final trained RF model. In the first approach, the prediction performance of the final trained RF model was compared to the prediction performance of the RF models developed in the previous works, as listed in Table 1.

Meanwhile, in the second approach, the prediction performance of the final trained RF model was compared to the prediction performance of several traditional empirical propagation models that are available in CloudRF, i.e., COST231, SUI, ECC-33 and ITM. The assessment and evaluation were carried out using the test datasets prepared earlier. Based on the literature review, COST231, SUI, ECC-33, and ITM have been widely used in wireless network planning, as listed in Table 5.

**TABLE 5.** List of previous works using propagation models COST231, SUI, ECC-33 and ITM.

| Traditional Empirical Propagation Model | List of Previous Works |
|---|---|
| COST231 | [3], [58]–[61] |
| SUI | [61]–[66] |
| ECC-33 | [3], [59], [62]–[64] |
| ITM | [67]–[69] |

### E. WEB-BASED ML MODEL DEVELOPMENT AND DEPLOYMENT

The development and deployment of MLOE is conducted according to the three stages described in Fig. 8.

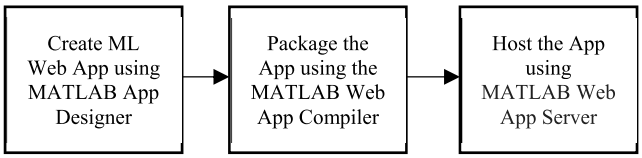


**FIGURE 8.** The development stages of the web app.

In the first stage, the Graphical User Interface (GUI) was designed into three main sections: (i) Features calculation and generation function; (ii) RSRP ML-based prediction model; and (iii) Features specification description and other relevant references. Using a MATLAB script file, the GUI was designed, programmed, and linked to the final trained RF model.

In the second stage, the GUI functions, data, and settings that define the final web application were compiled into a deployment format. In the third stage, the compiled project file was deployed in an application server with MATLAB Web App Server program and the appropriate MATLAB Runtime version. MATLAB Runtime is a collection of shared libraries and code that enables the packaged MATLAB applications to be utilized on a device without the MATLAB program installed. Therefore, the end users can access MLOE through a web browser using HTTP or HTTPS protocols. The source code for the development of MLOE is available publicly through an open-source GitHub repository at [70].

## IV. RESULTS AND DISCUSSIONS

### A. FEATURES IMPORTANCE

The training result of the RF model with all features equal to 3.23 dB and 0.90 for $\boldsymbol{RMSE}$ and $\boldsymbol{R^2}$, respectively. Meanwhile, the feature selection analysis results are shown in Table 6. All features used in this study influence the RSRP prediction process based on the results. However, the level of influence of each feature on the RSRP prediction process is varied. Therefore, the PDP analysis will reveal in detail the level of influence of these features.

In Fig. 9, the PDP revealed the feature Distance is negatively correlated to the response variable. This trend coincides with the domain knowledge, where the larger the separation distance between BS and UE, the weaker the received signal strength at the UE location. There is a considerable variation in the response variable for distance values varies from 0 to 1200 meters, contributing to 4.34 dBm SD. At 400 meters to 600 meters, the RSRP values will decrease drastically. However, the RSRP remained unchanged when the distance value reached 900 meters and above.

**TABLE 6.** Analysis results of the features selection process.

| No. | Features | $\boldsymbol{RMSE}$ (dB) | $\boldsymbol{R^2}$ |
|---|---|---|---|
| 1. | $D$ | 5.18 | 0.75 |
| 2. | $AZ_{offset}$ | 5.59 | 0.71 |
| 3. | $ELEV$ | 4.42 | 0.82 |
| 4. | $TILT_{offset}$ | 4.28 | 0.83 |
| 5. | $CLASS$ | 3.51 | 0.88 |
| 6. | $FQ$ | 3.48 | 0.89 |
| 7. | $OBS$ | 3.42 | 0.89 |

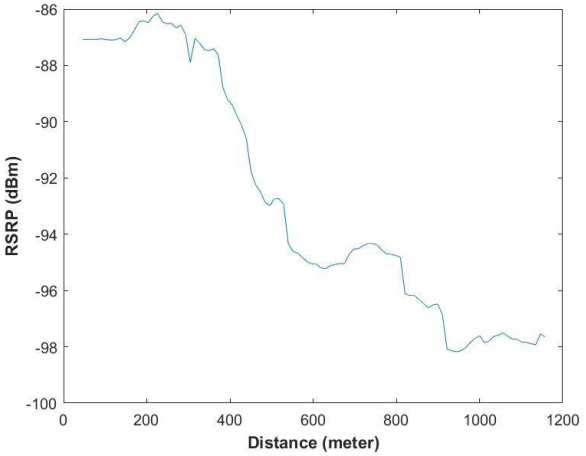


**FIGURE 9.** Partial dependence plot for distance.

The response variable is also negatively correlated with the $AZ_{offset}$ as shown in Fig. 10. This result is also coinciding with the principle of domain knowledge, where the received signal strength at the UE location gets weaker when its position is away from the boresight of the BS antenna. The variation of RSRP value is also substantial, especially when $AZ_{offset}$ in the range of 10 degrees to 20 degrees, and 50 degrees to 70 degrees. In the range of 20 degrees to 50 degrees, the rate of change in the RSRP value is only within 2 dB, while above 70 degrees, it was found that the rate of change in the RSRP value is very low, which is close to the zero value. Overall, the SD value for $AZ_{offset}$ is equal to 5.56 dBm, which is the highest compared to the other features applied in this study. Therefore, it is considered that the feature $AZ_{offset}$ is the most important and influential feature of this study.

In contrast with the PDP analysis result for $ELEV$ feature, as shown in Fig. 11. The graph slope trend showed a positively correlated, which means the closer UE is to the BS, the better signal strength will be received. This coincides with the principle of domain knowledge. A drastic change in the RSRP value can be seen in the range of 3 degrees to 7 degrees. In the

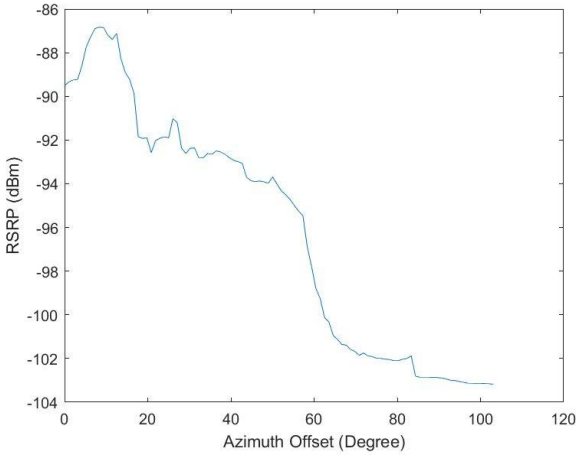


**FIGURE 10.** Partial dependence plot for azimuth offsets angle.

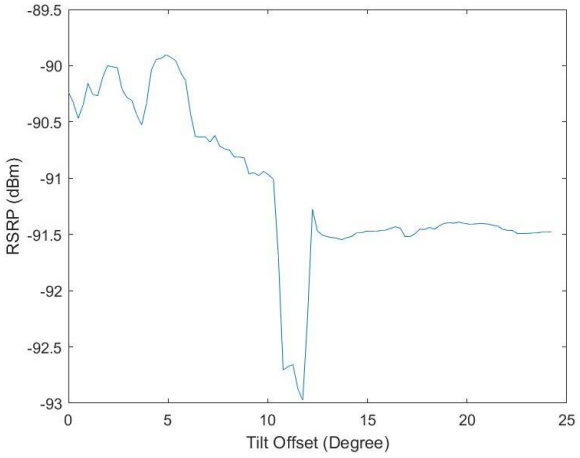


**FIGURE 12.** Partial dependence plot for tilt offset angle.

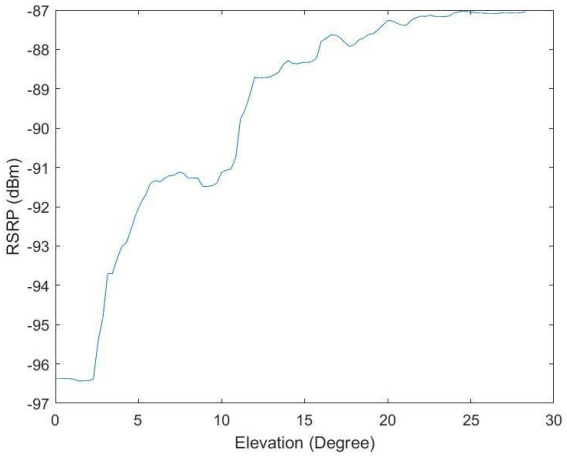


**FIGURE 11.** Partial dependence plot for elevation angle.

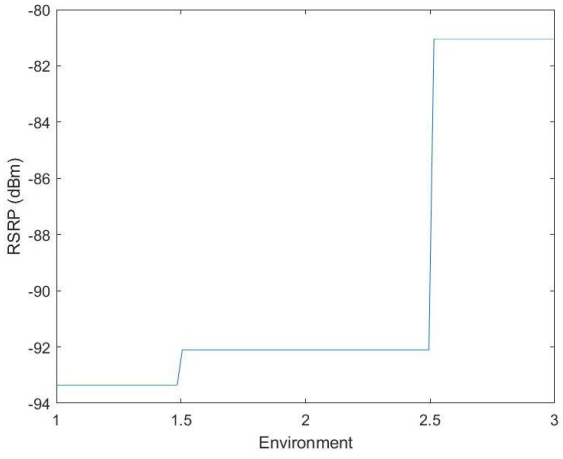


**FIGURE 13.** Partial dependence plot for environment category.

range of 11 degrees to 25 degrees, the RSRP values remain in an upward trend, within the range of 4dBm.

As for $TILT_{offset}$ feature, the graph trend is more towards a negative correlation. However, the graph slope (Fig. 12) is not too steep compared to the other feature, i.e., $D$, $AZ_{offset}$ and $ELEV$. This result coincides with the domain knowledge where the farther the UE location from the BS antenna boresight on the vertical plane, the received signal strength will get weaker. The SD for the $AZ_{offset}$ is 0.69 dBm, which means the influence level on the changes in RSRP value is not very high. Between 0 to 13 degrees, the variation of RSRP values is within 1 dBm. However, from 10 degrees to 13 degrees, the RSRP values decreased drastically within 2 dBm. This may be due to the reduction in the receiving signal strength when transitioning from the main lobe to the side lobe. After that, the RSRP values remain unchanged.

For environmental category features (refer to Fig. 13), a large gap in the RSRP value separates class 3 (open area) from class 2 (suburban) and class 1 (urban). The SD for this feature is equal to 5.03 dB, the second highest after the Azimuth Offset. So, it is proven that the type of environment is an essential and influential feature in the RSRP prediction process. This coincides with the domain knowledge principle, where signal propagation attenuation in open areas is much lower than in suburban and urban areas. However, for Class 2 and Class 1, the difference is not too significant. This is likely due to the less dense urban topology characteristics in Putrajaya. Therefore, more training dataset for a different type of urban topology is required to understand its influence better.

For the operating frequency features (refer to Fig. 14), it was found that the gap in the response variable during the RSRP forecasting process is relatively straightforward in differentiating between the 1800 MHz and 2600 MHz bands. The generated SD value of 1.33 dBm shows that the frequency bands feature is important and influential to the RSRP prediction process. PDP results show that the signal propagation performance on the lower frequency band is better, in line with the domain knowledge.

Based on the PDP analysis results in Fig. 15, the signal status features do not significantly contribute to the RSRP prediction process. The gap between LOS and NLOS categories

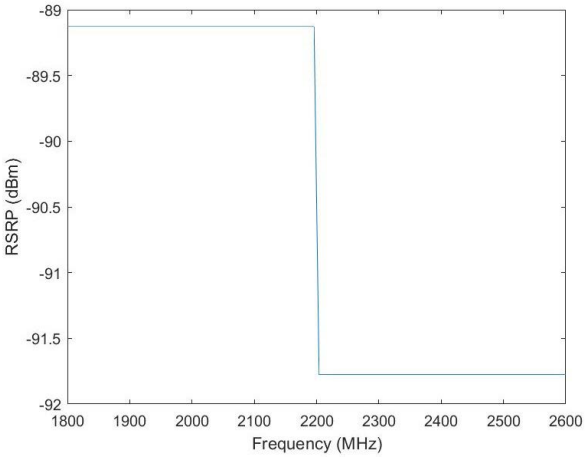


**FIGURE 14.** Partial dependence plot for frequency bands.

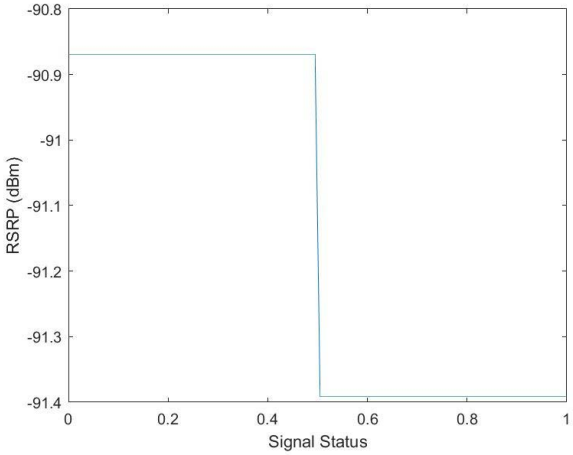


**FIGURE 15.** Partial dependence plot for signal status.

is around 0.5 dBm, contributing to 0.26 SD. Therefore, this feature is the least influential in this study. Perhaps due to the rough classifying of the signal status characteristics, which does not consider the difference in the number of obstacles and the die electric properties of each obstacle, the contribution of this parameter is seen as not very prominent compared to the contribution from the other features. However, it's still contributing to the final RSRP prediction process by a small margin.

Overall, it can be concluded that features applied in this study are influential based on priority order, as shown in Table 7.

### B. ML-BASED PREDICTION MODEL

In Fig. 16, the prediction performance of the final trained RF model was compared to the prediction performance of the RF model produced in previous works. It was found that the performance of the final trained RF model in this study is better than that of the RF model developed in previous works. After the hyperparameter tuning process, the final

**TABLE 7.** Priority order of influence features.

| Influence Priority | Features | Standard Deviation (dBm) |
|---|---|---|
| 1. | $AZ_{offset}$ | 5.56 |
| 2. | $CLASS$ | 5.03 |
| 3. | $D$ | 4.34 |
| 4. | $ELEV$ | 2.97 |
| 5. | $FQ$ | 1.33 |
| 6. | $TILT_{offset}$ | 0.69 |
| 7. | $OBS$ | 0.26 |

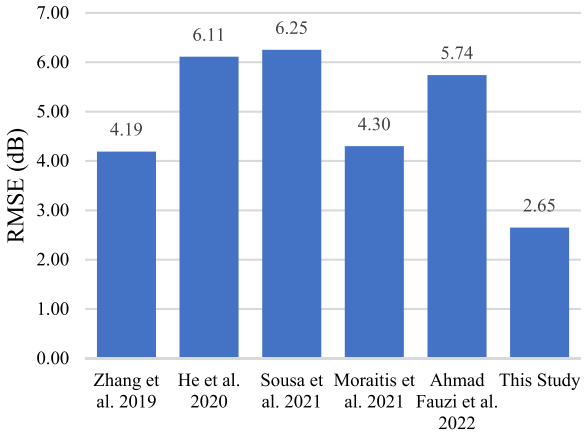


**FIGURE 16.** Comparison of final trained RF model performance with the RF model produced in the previous works.

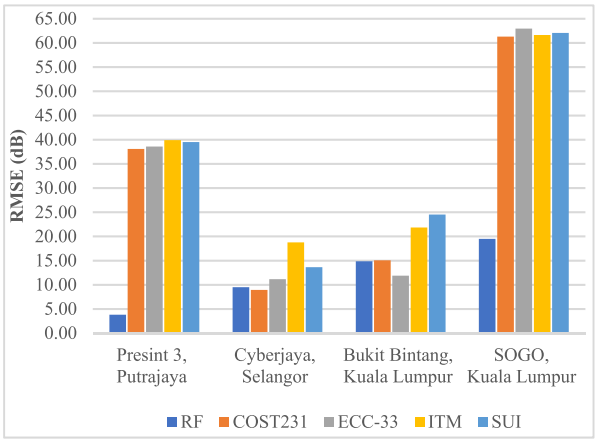


**FIGURE 17.** Comparison of final trained RF model performance with the traditional empirical propagation models using test dataset in an urban area.

trained RF model has achieved 2.65 dB and 0.93 of *RMSE* and $R^2$ respectively. This indicates that the features applied in this study are very suitable and influential for the mobile network path loss study. Compared to the features applied in the previous works, the combination of features in this study, especially for $D$, $AZ_{offset}$, $ELEV$ and $TILT_{offset}$ is capable of mapping more precisely the actual position of the UE in the main lobe beam of the BS antenna.

In Fig. 17 to Fig. 19, the prediction performance of the final trained RF model was compared to the prediction

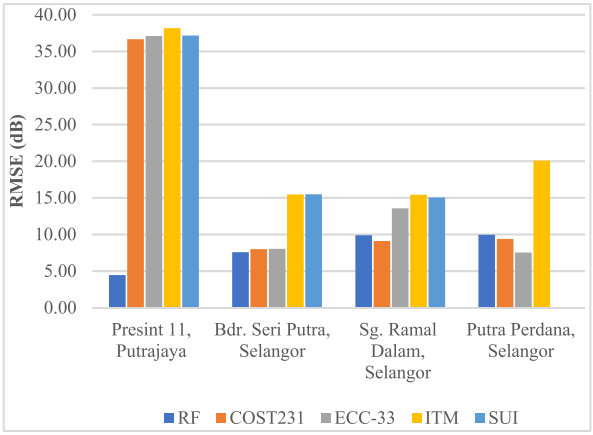


**FIGURE 18.** Comparison of final trained RF model performance with traditional empirical propagation models using test dataset at the suburban area.

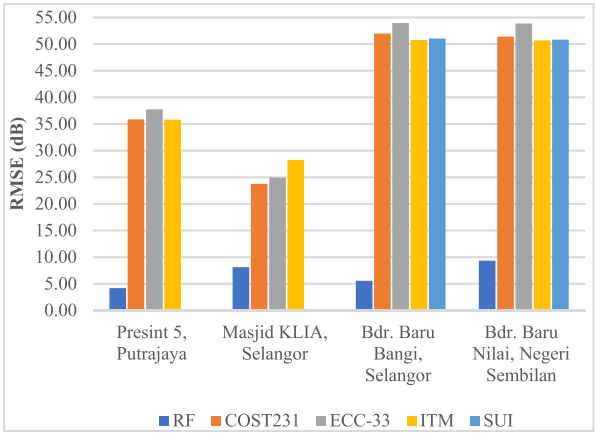


**FIGURE 19.** Comparison of final trained RF model performance with traditional empirical propagation models using test dataset in open areas.

performance of traditional empirical propagation models, i.e., COST231, SUI, ECC-33 and ITM using the test datasets.

Overall, the performance of the final trained RF prediction model showed better performance than the traditional empirical propagation model in urban areas (refer to Fig. 17). As expected, the *RMSE* of the final trained RF model was below 7 dB in the study area and increased when tested outside Putrajaya. However, the differences were not significant in the test area of Cyberjaya and Selangor. Meanwhile, the difference was almost three times for the test areas in Bukit Bintang and SOGO, which fall under the dense urban area category. The latter was expected because the landscape planning and urban design in both test areas differ from Putrajaya.

For the suburban test area (refer to Fig. 18), the performance of the final trained RF prediction model in the study area remains below 7dB. While the RF model prediction accuracy for all test locations outside Putrajaya surprisingly showed not more than 10 dB even though the build-up characteristics in some test areas were different. This may be due to structure height, density, and material properties in suburban areas, which are less significant than in urban areas. Meanwhile, it was found that the prediction performance of the final trained RF prediction model in suburban areas is more consistent than the traditional empirical propagation models. The traditional empirical propagation model tends to be inconsistent and less accurate, especially for SUI and ITM.

Likewise, in open test areas (refer to Fig. 19), the performance of the final trained RF prediction model inside the study area remains below 7 dB and outside the study area remains below 10 dB. This indicates that the performance of the RF prediction model is also better for open area categories compared to the traditional empirical propagation models.

In conclusion, this study's final trained RF prediction model proved to predict the RSRP value more accurately than the traditional empirical propagation models. Even outside the study area, especially in suburban and open areas, the final trained RF prediction model is seen as capable of providing reliable prediction capabilities.

Meanwhile, a more detailed training dataset is required for urban areas, especially for features that are more specifically described on differentiation types of urban design and signal propagation characteristics. Therefore, the ML algorithm will be able to distinguish more precisely the type and properties of the material that interact with the transmitted radio wave signal.

### C. WEB-BASED APPLICATION

Fig. 20 shows the final GUI of MLOE, which can be accessed openly through the Internet [16]. In the Calculator Function Section, users need to enter the BS technical specifications in Antenna/Tx Section, while the UE specifications are in Mobile Equipment/Rx Section. With the information entered, this calculator function will generate the output value of the features on the Features Value column. Thus, to perform the RSRP predictions process using the generated feature values, the user needs to press the Value Transfer button to send the generated values to the Prediction Model Section. However, the Prediction Model Section can only accept the value of the feature within the allowed range, which is equal to the IQR score as described earlier. If the value transferred is outside the allowed range, the value transfer process will be failed or be incomplete.

In general, the RSRP prediction process can be implemented in two approaches which are (i) predict the RSRP value by using the calculator functio; and (ii) predict the RSRP value by directly submitting the values of the required features in the Prediction Model Section. The GUI will directly interact with the final trained RF algorithm deployed in MATLAB Web APP Server to produce the RSRP prediction result.

In the Indicator and Description Section, there are three separate tabs. The first tab explains the feature characteristics and the related abbreviations. Meanwhile, the second tab explains information regarding RSRP signal strength categories. Finally, the third tab included background information regarding the ML algorithm and a disclaimer for user

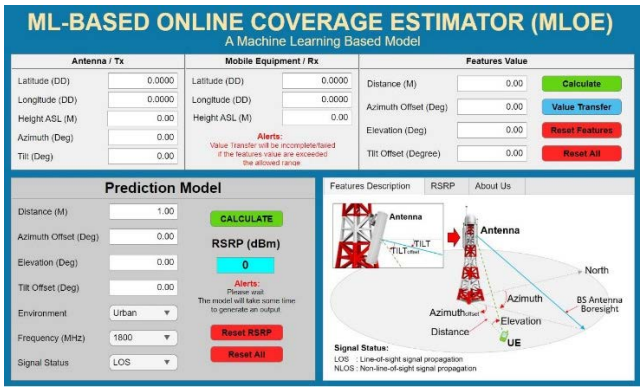


**FIGURE 20.** GUI of MLOE.

reference. At the same time, the author's e-mail address and YouTube page link for a tutorial video [71] were included for user' convenience.

Overall, MLOE can be utilised for mobile network planning before and after BS deployment. Before BS deployment, this model can identify the best location for BS to be deployed to achieve optimal coverage. Meanwhile, the model can be applied after BS deployment to perform antenna adjustment and fine-tuning, especially on the antenna azimuth and down-tilting angle.

## V. CONCLUSION

This article introduces and explains the concept and methodology of developing MLOE for predicting the coverage of 4G mobile networks through RSRP. In this study, the MATLAB App Designer was used to design and program the MLOE before being compiled and deployed into MATLAB Web App Server. The usage of seven unique features in this study has proven the capability to precisely describe the UE's actual location with reference to the BS antenna. As a result, it is enhanced the prediction accuracy to 2.65 dB and 0.93 for *RMSE* and $R^2$ respectively. The performance of MLOE is proven better than traditional empirical techniques and previous works. Finally, future work is to extend the MLOE function for 5G networks. Furthermore, the capabilities of remote sensing satellite image data will be exploited to produce more details and precise information related to the urban design dan characteristics of signal propagation. Besides, the benefits of MLOE can be further expanded by generating prediction results in spatial format, which allows for seamless integration into existing mobile network planning and monitoring systems.

## REFERENCES


[1] A. Yeong, "MCMC: Celcom, Digi and U Mobile failed to meet broadband service standards at several locations," Malay Mail, Malaysia, Tech. Rep., Jun. 2022.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 3GPP. *Advanced Plans for 5G*. Accessed: Dec. 29, 2021. [Online]. Available: https://www.3gpp.org/news-events/2210-advanced_5g

[7] E. Alimpertis, A. Markopoulou, C. Butts, and K. Psounis, "City-wide signal strength maps: Prediction with random forests," in *Proc. World Wide Web Conf. (WWW)*, 2019, pp. 2536–2542, doi: 10.1145/3308558.3313726.

[8] M. Behjati, M. A. Zulkifley, H. A. H. Alobaidy, R. Nordin, and N. F. Abdullah, "Reliable aerial mobile communications with RSRP & RSRQ prediction models for the Internet of Drones: A machine learning approach," *Sensors*, vol. 22, no. 15, p. 5522, Jul. 2022, doi: 10.3390/s22155522.

[9] 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.

[10] MCMC. (2020). *Pelan Jalinan Digital Negara (JENDELA)*. Malaysian Communications and Multimedia. Accessed: Aug. 18, 2022. [Online]. Available: https://myjendela.my/en-GB/

[11] M. F. A. Fauzi, *Machine Learning Training Dataset—Nemo Handy*. Accessed: Oct. 3, 2022. [Online]. Available: https://github.com/fazuwanfauzi/4G-LTE-RSRP-Prediction-Model-Ver2/tree/main/Dataset/Training

[12] 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.

[13] 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.

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

[15] M. F. Ahmad Fauzi, R. Nordin, N. F. Abdullah, and H. A. H. Alobaidy, "Mobile network coverage prediction based on supervised machine learning algorithms," *IEEE Access*, vol. 10, pp. 55782–55793, 2022, doi: 10.1109/ACCESS.2022.3176619.

[16] FKAB. *4G RSRP Prediction Model*. Accessed: Aug. 20, 2022. Universiti Kebangsaan Malaysia, [Online]. Available: https://MATLAB.ukm.my/webapps/home/

[17] 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.

[18] 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.

[19] 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, Jul. 2019, doi: 10.1109/MVT.2019.2921627.

[20] Y. Fu, K. N. Doan, and T. Q. S. Quek, "On recommendation-aware content caching for 6G: An artificial intelligence and optimization empowered paradigm," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 304–311, Aug. 2020, doi: 10.1016/j.dcan.2020.06.005.

[21] C. G and J. M. Roogi, "A quick review of ML algorithms," in *Proc. 6th Int. Conf. Commun. Electron. Syst. (ICCES)*, Jul. 2021, doi: 10.1109/ICCES51350.2021.9488982.

[22] M. Wu, "Financial transaction forecasting using neural network and Bayesian optimization," in *Proc. IEEE Int. Conf. Comput. Sci., Electron. Inf. Eng. Intell. Control Technol. (CEI)*, Sep. 2021, pp. 193–196, doi: 10.1109/CEI52496.2021.9574535.

[23] H. A. H. Alobaidy, M. Jit Singh, M. Behjati, R. Nordin, and N. F. Abdullah, "Wireless transmissions, propagation and channel modelling for IoT technologies: Applications and challenges," *IEEE Access*, vol. 10, pp. 24095–24131, 2022, doi: 10.1109/ACCESS.2022.3151967.
[24] K. Tan, D. Bremner, J. Le Kernec, L. Zhang, and M. Imran, "Machine learning in vehicular networking: An overview," *Digit. Commun. Netw.*, vol. 8, no. 1, pp. 18–24, Feb. 2022, doi: 10.1016/j.dcan.2021.10.007.
[25] 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.
[26] S. Alraih, I. Shayea, M. Behjati, R. Nordin, N. F. Abdullah, A. Abu-Samah, and D. Nandi, "Revolution or evolution? Technical requirements and considerations towards 6G mobile communications," *Sensors*, vol. 22, no. 3, p. 762, Jan. 2022, doi: 10.3390/s22030762.
[27] M. Behjati, M. H. Mazlan, A. M. Ramly, R. Nordin, and M. Ismail, "What is the value of limited feedback for next generation of cellular systems?" *Wireless Pers. Commun.*, vol. 110, no. 3, pp. 1127–1142, Feb. 2020, doi: 10.1007/s11277-019-06777-1.
[28] N. Ye, J. Yu, A. Wang, and R. Zhang, "Help from space: Grant-free massive access for satellite-based IoT in the 6G era," *Digit. Commun. Netw.*, vol. 8, no. 2, pp. 215–224, Apr. 2022, doi: 10.1016/j.dcan.2021.07.008.
[29] Y. Chen, W. Liu, Z. Niu, Z. Feng, Q. Hu, and T. Jiang, "Pervasive intelligent endogenous 6G wireless systems: Prospects, theories and key technologies," *Digit. Commun. Netw.*, vol. 6, no. 3, pp. 312–320, Aug. 2020, doi: 10.1016/j.dcan.2020.07.002.
[30] 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.
[31] X. Zhang, *LTE Optimization Engineering Handbook*. Singapore: Wiley, 2017.
[32] (2022). *CloudRF*. [Online]. Available: https://cloudrf.com/
[33] T. N. B. Duong and N. Q. Sang, "Distributed machine learning on IAAS clouds," in *Proc. 5th IEEE Int. Conf. Cloud Comput. Intell. Syst. (CCIS)*, Nov. 2018, pp. 58–62, doi: 10.1109/CCIS.2018.8691150.
[34] 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.
[35] D. Fernandes, G. Soares, D. Clemente, R. Cortesao, P. Sebastiao, F. Cercas, R. Dinis, and L. S. Ferreira, "Combining measurements and propagation models for estimation of coverage in wireless networks," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5, doi: 10.1109/VTCFall.2019.8891451.
[36] H. Cao, J. Finer, M. Megjhani, D. C. Nametz, V. Lorenzi, L. Mamykina, R. Meyers, S. C. Rossetti, and S. Park, "Machine learning model deployment using real-time physiological monitoring: Use case of detecting delayed cerebral ischemia," in *Proc. IEEE Healthcare Innov. Point Care Technol. (HI-POCT)*, Mar. 2022, pp. 42–45, doi: 10.1109/HI-POCT54491.2022.9744076.
[37] A. B. Kolltveit and J. Li, "Operationalizing machine learning models—A systematic literature review," in *Proc. IEEE/ACM 1st Int. Workshop Softw. Eng. Responsible Artif. Intell. (SERAI)*, May 2022, pp. 1–8. [Online]. Available: https://ieeexplore.ieee.org/document/9808768
[38] L. Baier and S. Seebacher, "Challenges in the deployment and operation of machine learning in practice," in *Proc. 27th Eur. Conf. Inf. Syst.*, May 2019, pp. 1–15.
[39] N. F. Abdullah and M. L. Hakim 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.
[40] H. A. H. Alobaidy, M. J. Singh, 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.*, vol. 9, no. 13, pp. 11614–11632, Jul. 2022, doi: 10.1109/jiot.2021.3131160.
[41] M. F. A. Fauzi. *Location of Tested BS and Distributions of Test Point*. Accessed: Oct. 3, 2022. [Online]. Available: https://github.com/fazuwanfauzi/4G-LTE-RSRP-Prediction-Model-Ver2/tree/main/Test Sites
[42] Keysight Technologies. *Nemo Handy*. Accessed: Aug. 18, 2022. [Online]. Available: https://www.keysight.com/my/en/product/NTH00000B/nemo-handy-handheld-measurement-solution.html#:~:text=NemoHandyisanAndroid,of-experience(QoE)
[43] J. Obuhuma, H. Okoyo, and S. McOyowo, "Real-time driver advisory model: Intelligent transportation systems," in *Proc. IST-Africa Week Conf. (IST-Africa)*, 2018, pp. 1–11.
[44] M. Hysenaj and E. Hoxha, "Spatio-textual technology: The future of web search," in *Proc. Int. Conf. Interact. Mob. Commun. Technol. Learn. (IMCL)*, Nov. 2015, pp. 210–213, doi: 10.1109/IMCTL.2015.7359588.
[45] N. Xu, S. Li, C. S. Charollais, A. Burg, and A. Schumacher, "Machine learning based outdoor localization using the RSSI of multibeam antennas," in *Proc. IEEE Workshop Signal Process. Syst. (SiPS)*, Oct. 2020, pp. 1–5, doi: 10.1109/SiPS50750.2020.9195235.
[46] H. Patel, "Beam refinement and beam tracking using Machine Learning Techniques in 5G NR RAN," M.S. thesis, Dept. Comput., Blekinge Inst. Technol., Karlskrona, Sweden, Jan. 2021.
[47] B. Szyk. *Azimuth Calculator*. Omni Calculator. Accessed: Aug. 14, 2022. [Online]. Available: https://www.omnicalculator.com/other/azimuth#azimuth-formula
[48] S. Mohammadjafari, S. Roginsky, E. Kavurmacioglu, M. Cevik, J. Ethier, and A. B. Bener, "Machine learning-based radio coverage prediction in urban environments," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 4, pp. 2117–2130, Dec. 2020, doi: 10.1109/TNSM.2020.3035442.
[49] B. Turan and S. Coleri, "Machine learning based channel modeling for vehicular visible light communication," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 9659–9672, Oct. 2021, doi: 10.1109/TVT.2021.3107835.
[50] H. A. H. Alobaidy, R. Nordin, M. J. Singh, N. F. Abdullah, A. Haniz, K. Ishizu, T. Matsumura, F. Kojima, and N. Ramli, "Low-Altitude-Platform-Based airborne IoT network (LAP-AIN) for water quality monitoring in harsh tropical environment," *IEEE Internet Things J.*, vol. 9, no. 20, pp. 20034–20054, Oct. 2022, doi: 10.1109/JIOT.2022.3171294.
[51] OSMF. (2022). *OpenStreetMap*. [Online]. Available: https://www.openstreetmap.org
[52] H. Elsherbiny, A. M. Nagib, H. Abou-zeid, H. M. Abbas, and H. S. Hassanein, "4G LTE Network Data Collection and Analysis along Public Transportation Routes," in *Proc. IEEE Global Commun. Conf.*, Dec. 2020, pp. 1–6.
[53] M. F. A. Fauzi. (2022). *Machine Learning Training Dataset—IQR Analysis*. [Online]. Available: https://github.com/fazuwanfauzi/4G-LTE-RSRP-Prediction-Model-Ver2/tree/main/IQR
[54] C. Molnar, *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*. Canada: Lean Publishing, 2019, p. 247.
[55] E. Zhu, J. Zhang, J. Yan, K. Chen, and C. Gao, "N-gram MalGAN: Evading machine learning detection via feature N-gram," *Digit. Commun. Netw.*, vol. 8, no. 4, pp. 485–491, Aug. 2022, doi: 10.1016/j.dcan.2021.11.007.
[56] F. Afshar, S. Seyedabrishami, and S. Moridpour, "Application of extremely randomised trees for exploring influential factors on variant crash severity data," *Sci. Rep.*, vol. 12, no. 1, p. 11476, Jul. 2022, doi: 10.1038/s41598-022-15693-7.
[57] MATLAB. (2022). Standard Deviation. The MathWorks. Accessed: Aug. 20, 2022. [Online]. Available: https://www.mathworks.com/help/MATLAB/ref/std.html#:~: text=std is operating.-,S%3Dstd(A%2Cw%2C%22all%22),versionsR2018bandlater.&text=S%3Dstd(A%2Cw%2Cdim)returnsthe,standarddeviationalongdimensiondim
[58] J. Liu, Y. Liu, Y. Li, G. Zhang, W. He, and H. Li, "Wireless network coverage simulation, measurement, and propagation model correction," in *Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, Dec. 2020, pp. 1004–1012, doi: 10.1109/ITAIC49862.2020.9339161.
[59] P. K. Sharma, D. Sharma, and T. V. Sai, "Optimization of propagation path loss model in 4G wireless communication systems," in *Proc. 2nd Int. Conf. Inventive Syst. Control (ICISC)*, Jan. 2018, pp. 1245–1248, doi: 10.1109/ICISC.2018.8399004.
[60] Z. Waheed, U. R. Kamboh, M. N. Shehzad, M. D. Taqdees, M. Usman, and A. Fatima, "Measurements of deterministic propagation models through field assessments for long-term evaluation," in *Proc. Int. Conf. Innov. Comput. (ICIC)*, Nov. 2021, pp. 1–6, doi: 10.1109/ICIC53490.2021.9693000.
[61] J. Zhang, L. Xu, R. Zhang, W. He, and Y. Wang, "Atoll-based propagation model correction and actual measurement," in *Proc. IEEE 4th Adv. Inf. Technol., Electron. Autom. Control Conf. (IAEAC)*, Dec. 2019, pp. 662–666, doi: 10.1109/IAEAC47372.2019.8997802.
[62] A. A. P. D. Carvalho, I. S. Batalha, M. A. Neto, B. L. Castro, F. J. B. Barros, J. P. L. Araujo, and G. P. S. Cavalcante, "Adjusting large-scale propagation models for the Amazon region using bioinspired algorithms at 1.8 and 2.6 GHz frequencies," *J. Microw., Optoelectronics Electromagn. Appl.*, vol. 20, no. 3, pp. 445–463, Sep. 2021, doi: 10.1590/2179-10742021v20i31099.

[63] B. J. Cavalcanti, G. A. Cavalcante, M. T. G. Santos, G. M. Cantanhede, A. G. D'Assuncäo, and L. M. de Mendonça, "Propagation models tuning for long term evolution and long term evolution-advanced using genetic algorithms and least mean square," in *Proc. 12th Eur. Conf. Antennas Propag. (EuCAP)*, no. CP741, 2018, p. 5, doi: 10.1049/cp.2018.1074.

[64] B. J. Cavalcanti and L. M. de Mendonca, "Analysis of empirical propagation models in suburban areas at 800 MHz and 1.8 GHz," in *Proc. IEEE Int. Symp. Antennas Propag. USNC-URSI Radio Sci. Meeting (APS/URSI)*, Dec. 2021, pp. 225–226, doi: 10.1109/APS/URSI47566.2021.9704513.

[65] S. O. Hasan and S. S. Abdullah, "Path loss estimation for some korek-telecom sites operating at (1.8) GHz and (2.1) GHz for urban and sub-urban area in erbil city," *Adv. Sci., Technol. Eng. Syst. J.*, vol. 5, no. 5, pp. 869–875, 2020, doi: 10.25046/AJ0505106.

[66] M. Pinem, S. I. Rezkika, T. F. Astarani, A. P. Narulitasari, and S. C. Y. Pardede, "Rooftop to street radio propagation analysis on GSM 1800 frequency spectrum," in *Proc. 5th Int. Conf. Electr., Telecommun. Comput. Eng. (ELTICOM)*, Sep. 2021, pp. 107–111, doi: 10.1109/ELTICOM53303.2021.9590154.

[67] A. Bjelopera, E. Dumic, and M. Kajinic, "Simulation of radio wave propagation models on 800 MHz and 1.8 GHz in the city of dubrovnik," in *Proc. 2nd Eur. Conf. Electr. Eng. Comput. Sci. (EECS)*, Dec. 2018, pp. 510–515, doi: 10.1109/EECS.2018.00099.

[68] 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.

[69] J. Jang, L. Xu, B. Yee, E. Petsalis, S. Beck, V. I. Lang, and H. Chew, "Propagation models: Large scale and site-specific," in *Proc. IEEE Int. Symp. Antennas Propag. USNC/URSI Nat. Radio Sci. Meeting*, Jul. 2018, pp. 93–94, doi: 10.1109/APUSNCURSINRSM.2018.8608578.

[70] M. F. A. Fauzi. *4G RSRP Prediction Model Source Code*. UKM. Accessed: Oct. 11, 2022. [Online]. Available: https://github.com/fazuwanfauzi/4G-LTE-RSRP-Prediction-Model-Ver2/tree/main/Source Code

[71] FKAB. Tutorial Video: 4G RSRP Prediction Model. Universiti Kebangsaan Malaysia. Accessed: Oct. 3, 2022. [Online]. Available: https://www.youtube.com/watch?v=G9GQFx6ePKE

**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 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 Universiti Kebangsaan Malaysia, in 2001, and the Ph.D. degree from the University of Bristol, U.K., in 2011. He is currently a Professor at 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 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 Universiti Kebangsaan Malaysia, Selangor, Malaysia. Her research interests include 5G, millimeter waves, LTE-A, vehicular networks, massive MIMO, space-time coding, fountain codes, 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, and the Ph.D. degree in electrical, electronic and systems engineering from the Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), in 2022. His research interests include wireless sensor networks, wireless communication, the IoT, and channel propagation modeling and estimation.

**MEHRAN BEHJATI** received the B.Sc. degree in electrical and electronic engineering from Islamic Azad University, Iran, in 2009, and the M.Sc. and Ph.D. degrees in communication engineering from the National University of Malaysia (UKM), in 2013 and 2017, respectively. He is currently a Postdoctoral Researcher with the Department of Electrical, Electronics, and System Engineering, UKM. His research interests include advanced multiple-input multiple-output (MIMO) systems, interference management, limited-feedback techniques, aerial wireless communications, the IoT, the Internet of Drone (IoD), and green communications.