

Article

Improving Path Loss Prediction Using Environmental Feature Extraction from Satellite Images: Hand-Crafted vs. Convolutional Neural Network

Usman Sammani Sani ¹ , Owais Ahmed Malik ^{1,2}  and Daphne Teck Ching Lai ^{1,*} 

¹ School of Digital Science, Universiti Brunei Darussalam, Tungku Link, Gadong BE1410, Brunei; 19h8338@ubd.edu.bn (U.S.); owais.malik@ubd.edu.bn (O.A.M.)

² Institute of Applied Data Analytics, Universiti Brunei Darussalam, Tungku Link, Gadong BE1410, Brunei

* Correspondence: daphne.lai@ubd.edu.bn

Abstract: There is an increased exploration of the potential of wireless communication networks in the automation of daily human tasks via the Internet of Things. Such implementations are only possible with the proper design of networks. Path loss prediction is a key factor in the design of networks with parameters such as cell radius, antenna heights, and the number of cell sites that can be set. As path loss is affected by the environment, satellite images of network locations are used in developing path loss prediction models such that environmental effects are captured. We developed a path loss model based on the Extreme Gradient Boosting (XGBoost) algorithm, whose inputs are numeric (non-image) features that influence path loss and features extracted from images composed of four tiled satellite images of points along the transmitter to receiver path. The model can predict path loss for multiple frequencies, antenna heights, and environments such that it can be incorporated into Radio Planning Tools. Various feature extraction methods that included CNN and hand-crafted and their combinations were applied to the images in order to determine the best input features, which, when combined with non-image features, will result in the best XGBoost model. Although hand-crafted features have the advantage of not requiring a large volume of data as no training is involved in them, they failed in this application as their use led to a reduction in accuracy. However, the best model was obtained when image features extracted using CNN and GLCM were combined with the non-image features, resulting in an RMSE improvement of 9.4272% against a model with non-image features only without satellite images. The XGBoost model performed better than Random Forest (RF), Extreme Learning Trees (ET), Gradient Boosting, and K Nearest Neighbor (KNN) based on the combination of CNN, GLCM, and non-image features. Further analysis using the Shapley Additive Explanations (SHAP) revealed that features extracted from the satellite images using CNN had the highest contribution toward the XGBoost model's output. The variation in values of features with output path loss values was presented using SHAP summary plots. Interactions were also observed between some features based on their dependence plots from the computed SHAP values. This information, when further explored, could serve as the basis for the development of an explainable/glass box path loss model.

Keywords: path loss; hand-crafted; satellite image; feature extraction; environment



Citation: Sani, U.S.; Malik, O.A.; Lai, D.T.C. Improving Path Loss Prediction Using Environmental Feature Extraction from Satellite Images: Hand-Crafted vs. Convolutional Neural Network. *Appl. Sci.* **2022**, *12*, 7685. <https://doi.org/10.3390/app12157685>

Academic Editor: José Salvador Sánchez Garreta

Received: 8 June 2022

Accepted: 27 July 2022

Published: 30 July 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Accurate path loss prediction is important in wireless networks. Power distribution within the coverage area can be accurately ascertained during Link Budget, leading to a network with a better quality of service if the predicted values are effectively used during design, expansion, or maintenance [1–3]. Deterministic path loss models provide a reasonable accuracy but are computationally expensive and require details that include a 3D model of the environment, building materials, and so on. Empirical models are much easier to use but are prone to errors. Meanwhile, machine learning (ML) models are more

accurate than empirical models and sometimes exceed the accuracy of deterministic models. Such ML models were developed with algorithms such as Multiple Layer Perceptron (MLP), Support Vector Regression (SVR), Random Forest (RF), and XGBoost [4–6]. Furthermore, a Convolutional Neural Network (CNN) is employed to extract features from images that describe the effects of environment due to the dependence of path loss on the propagation path/environment. Such images include satellite images for a wide area [7], pointwise satellite images [8], segmented images with shadings that represent heights of buildings in an urban environment [9], segmented images based on some parameters, with building height inclusive [10], and pseudoimages representing the fusion of various data types [11].

Most of the models developed are for fixed frequency, antenna height, and environment. However, since path loss varies with the environment, such model prediction will be inaccurate in other scenarios. Considering the fact that path loss prediction is performed before networks are deployed, it is reasonable to develop models that can be used to make accurate predictions for varieties of environments and network parameters. Otherwise, the model will only be suitable for the existing network it was designed for or a network with almost the same parameters in a similar environment. Although ML models make better predictions, empirical models have the advantage of being versatile as they consider a range of frequencies, distances, and antenna heights over which these models are valid. Some of these models have parameters tailored to a specific type of environment, such as COST 231 Hata [12], and Stanford University Interim (SUI) model [13]. Therefore, there is a need to have machine learning path loss models valid for multiple parameters and environments so that they can provide accurate predictions during design and prior to network deployment. This will eliminate the cost of taking measurements under certain network parameters from individual environments and developing individual models specific to the environments and selected network parameters.

Ayadi et al. [14] developed an MLP model for the prediction of a number of frequencies in the Ultra-High-Frequency (UHF) band. The study considered the fact that the propagation path from the transmitter to receiver position varies such that some portions can be classified as either rural, urban, or suburban environments. Thus, distances covered in each type of environment were included as features for model development. The problem with such a model is that it does not capture the micro variations of the environment types along the propagation path; rather, it is developed based on distances covered in each environment type without a detailed description. Nguyen et al. [15] developed a model valid at frequencies between 0.8 and 70 GHz. The model was developed based on the features distance, frequency, propagation scenario, and environment type. The propagation scenario could be either “above rooftop” or “below rooftop”. Above rooftop is a scenario in which one of the antenna heights is above the surrounding rooftops, and below rooftop is when all antenna heights are below the surrounding rooftops. The model does not capture environment details, unlike that in [16], where features extracted from satellite images are included during model development to capture environmental details.

Sani et al. [17] also developed models for multiple frequencies, multiple antenna heights, and multiple environments. A clutter height feature was used in representing the various types of environments such that each type of environment had an integer value representing it. The problem with this method is that it does not take into consideration the micro differences within each type of environment, whereas [16] included satellite images to acknowledge the effects of these micro variations. The ResNet50V2 used in [16] takes time and memory space to train. This work proposes the use of a less complex CNN that will train faster and occupy less memory space.

Another issue to note is the dependence of a CNN’s accuracy on a large volume of data and a large amount of time it takes a CNN to get trained with the data, even though the prediction time is minimal. An alternative is to use hand-crafted feature extraction methods to deal with data limitation as well as the large training time. Sotiroudis et al. used the Segmentation-based Fractal Texture Analysis (SFTA) texture extraction method to extract features from images with grayscale shadings of building heights between the transmitter

and receiver and developed three models with features based on RF, Extreme Gradient Boosting Machine (XGBoost), and Light Gradient Boosting Machine (LightGBM) [18]. However, a comparison between the performance of SFTA and CNN was not performed. To the best of our knowledge, their work is the first that incorporated texture feature extraction in a path loss problem. Predictions from a LightGBM model were more accurate, although the hyper-parameter optimization that was carried out in this work was not carried out in theirs to ascertain the optimal hyper-parameters of the algorithms. In other domains, comparisons were made between the performance of hand-crafted features and CNN, in which the CNN-extracted features performed better [19,20]. Combining CNN features and hand-crafted features improved the accuracy [21,22].

The aim of this work is to develop a single ML-based path loss model valid for some network parameters (UHF band and antenna heights) and environments (rural, suburban, urban, and urban highrise) with the effect of terrain/elevation taken into consideration. The essence is to have the ML path loss models valid for various values of network parameters and environment types, similar to existing empirical models. The ability of ML algorithms to model non-linear relationships improves performance over the existing empirical models. Numeric features combined with satellite image features were used as input to the model. Satellite images are used to capture micro and macro variations within the environments, unlike in existing empirical models, where such variations are not captured, resulting in poor predictions in new environments. The performance of features extracted using various methods (hand-crafted and CNN) and their combinations on the model's output was evaluated. Hand-crafted features were investigated to ascertain the possibility of their use in this application, knowing that they need no training, unlike CNN, which requires training with a lot of data. However, since combining hand-crafted and CNN features was reported to improve accuracy in some applications, we also investigated it, and improvements in accuracy and training time were observed compared to [16]. In addition, the study explores feature importance and analysis. Feature importance was not conducted in all works cited in this paper, except in [17], where only non-image features were considered for model development. Our contributions are:

- Review of path loss prediction based on image processing using CNN. Most works presented ML-based image processing approaches for path loss prediction, but to the best of our knowledge, no such comprehensive review of such existing works was supplied.
- Investigation on the use of various hand-crafted feature extraction techniques and their combinations on 2D satellite images in path loss prediction.
- Combining CNN features with hand-crafted features to improve the accuracy of path loss prediction. While this has been applied in computer vision tasks, to the best of our knowledge, this has not been applied to the problem of path loss prediction.
- Interpretation of model prediction based on Shapley Additive Explanations (SHAP). Most ML-based path loss prediction works do not analyze the features deeply, including the comparison of different types of features.

2. Related Works

The effects of the environment on path loss motivated the incorporation of environment details in the form of 2D satellite images into path loss modeling. A number of works have emerged in which images are processed using CNN. Various image types are being processed that include satellite images, segmented images based on some parameters, pseudoimages, and combinations of various images.

2.1. Satellite Images

Models that take satellite images as input was developed by scholars by either training a customized CNN or by the use of pre-trained CNN models such as VGG-16, ResNet-50, and ResNet50V2 via transfer learning as follows.

2.1.1. Training a Customized CNN with Satellite Images

Thrane et al. [8] predicted path loss in a campus environment (suburban) from a model that integrated satellite image input with other numerical data features. The satellite images are passed through five convolutional layers, and the numerical data are passed through fully connected layers. The outputs of the two networks are concatenated and passed onto an output neural network composed of two fully connected layers and a single node output layer. The data used in training and testing were measurements at 811 and 2630 MHz, with a fixed transmitting antenna height of 30 m and receiving antenna height of 1.5 m. The deep learning model performed better than 3D Ray Tracing and the empirical model. A similar architecture was developed to predict signal strength at 811 and 2630 MHz, but with reduced complexity in the CNN, whereby four convolutional layers were used with a reduced number of filters. A modification was also made to the numerical features used, such that latitude and longitude were removed, the frequency was included, and data from different environments were considered [23]. The model was upgraded in [24] by the addition of the receiver position side view image. Nguyen et al. modified the model in [23] such that the model predicts signal strength corrections of a 3GPP empirical model at 3.5 GHz. The final signal strength is obtained by the addition of the corrections to the signal strength computed using the 3GPP empirical model [25]. However, the image of the receiver position is not sufficient to describe the path of the signal from the transmitter to the receiver position.

2.1.2. Transfer Learning with Satellite Images

In order to train a model with more information about the path from transmitter to receiver location, Sani et al. [16] merged the satellite images of four locations along the path from the transmitter to receiver location into single images and extracted features from the images using a ResNet50V2. The satellite images were also from various environments, as in [23]. The extracted features were merged with eight other features and used for training and testing various regression algorithms. As ResNet50V2 takes a long period of time to train and requires large memory space, a customized CNN with lower complexity should have been used for the purpose.

Ates et al. [26] used VGG-16 and ResNet-50 to predict a path loss exponent and shadowing for 1.8 km² coverage areas using satellite images. Separate models were trained for the path loss exponent and shadowing values resulting from a 3D ray Tracing simulation at a frequency of 900 MHz, a transmitting antenna height of 300 m, and a receiving antenna height of 1.5 m. VGG-16 and ResNet-50 had the same accuracy in path loss exponent prediction, but ResNet-50 was better in terms of shadowing prediction by 1%. A modification of the two CNNs by including a parallel, fully connected layer to learn a histogram of the difference between path loss values and a reference path loss improved their performance by 2%. The process of learning two parameters from a single CNN is called multi-tasking. The prediction of the path loss exponent and shadowing has drawbacks. Both parameters were determined based on curve fitting/least squares regression, and these values contain errors and are not actual values. Training the CNN has its own error as well, thereby compounding the error of the curve fitting/least squares regression. In addition, separate models have to be developed for other frequency bands (for example, 1800 and 2600 MHz) for the same environment. Ahmadien et al. trained a VGG-16 to predict the histogram of path loss. Three models were developed at different single frequency values (900 MHz and 3.5 GHz) and antenna heights (80 and 300 m) [7]. The prediction of the path loss histogram does not give path loss values location-wise; rather, it gives the ranges of path loss values for a large area. Such information cannot be used in the optimization of cell radius and number of cells during design. A single model should have been developed for the various frequencies and antenna heights. This will save the time needed to train several models and memory allocation if the model is to be implemented in software.

2.2. Segmented Images

Images segmented based on building heights and/or terrain information have been used in path loss model development. Customized CNNs or modified forms of UNet were used for the purpose. The following subsections account for path loss models developed based on segmented images.

2.2.1. Training a Customized CNN with Segmented Images

Lee et al. trained a CNN with 3D Ray tracing simulated data at 28 GHz to predict path loss exponent for 1 km² area based on an image input resulting from the combination of two images. The first image is a segmented image that presents the heights of buildings with respect to the height of the tallest building as shades of red, and the second image is a height above sea level difference between the transmitting antenna and ground level presented in shades of green [10]. However, building height information is available for only a few environments, making the model useless where building height information is unavailable. Though the model is for an urban environment, the effect of vegetation that is present in most urban environments is omitted, especially for a millimeter wave of 28 GHz, which is easily degraded and blocked. Satellite images should have been used to capture such effects, and the model is made versatile. In [9], images containing gray shadings based on building heights between transmitter and receiver were used in training a CNN for path loss prediction at 900 MHz, in which the training data were obtained using 3D Ray Tracing. The performance was comparable to that of a two hidden layer MLP trained with 23 features describing the path from the transmitter to the receiver. The model, however, has to be modified for use at frequencies other than 900 MHz.

Cheng et al. trained an Attention-Enhanced CNN to predict path loss within a sub-urban environment at 28 GHz. The model's input is an image segmented based on the Distance Embedded Local Area Multi-scanning (DE-LAMS) algorithm and has an accuracy better than that of a normal CNN, dilated CNN, and a Global Context (GC) CNN [27]. Kim et al. developed a 5G millimeter wave signal strength prediction model based on a 3D CNN, whose inputs are 3D images generated by the 3D-LAMS algorithm based on building and terrain information. Pooling layers in the CNN were avoided so as to prevent information loss. The model, however, neglects the effect of vegetation, which is a very important determinant of path loss/received signal strength and the millimeter frequency/frequencies considered during model development were not specified [28]. Wu et al. developed a path loss prediction model with numerical input characterizing network parameters and image-based inputs characterizing the environment. The images were developed based on grayscale encoding such that each type of land cover between the transmitter and receiver had its own gray level/shade. A CNN was used for image feature extraction, and a fully connected neural network was used for the numeric data. The outputs of the two networks are fused together and passed onto fully connected layers, the output of which is the predicted path loss. The model's accuracy was better than when a combination of vector-based environment characterization and numeric inputs or when the three input types were combined [29]. The encoding method used does not capture micro variations within each type of environment.

2.2.2. Training Modified UNet Architecture with Segmented Images

Ratnam et al. modified a UNet and trained it to predict an image showing path loss distribution within an area. The training was performed with 3D Ray Tracing data at 28 GHz over areas that are 512 (m²), with the transmitting antenna at the center. Inputs to the network are terrain information, building information, and foliage information for the area under study encoded as colored images. The inclusion of Line of Sight (LOS) data further improved prediction accuracy, which surpassed distance dependence prediction and conditional least squares. The model is named FadeNet, and it demonstrated faster prediction than 3D Ray Tracing [30].

Zhang et al. trained a UNet model named PLNet. The inputs are images representing the distribution of parameters, clutter, terrain, building, frequency, antenna height, etc., within the region of interest, and the output is an image showing path loss distribution. Predictions by PLNet outperformed those of empirical models and 3D Ray Tracing [31]. The authors did not specify the range of parameter values for which the model is valid. The models in [30,31] are based on UNet, and the output is a segmented image of the path loss distribution. Such a model can only be used for analysis rather than design as numeric path loss values are required.

2.3. Combination of Images

Hayashi et al. used the feature extraction part of AlexNet to process images before passing the images together with numeric features to six fully connected layers. The image inputs include aerial maps, building occupancy maps, and building height maps. For each image type, either image at the receiver location alone or at both transmitter and receiver locations were considered, with each image having a separate CNN for processing. The results showed that for each input type, using images at both transmitter and receiver locations was better and that using the building height map as the input had the highest accuracy [32]. In a similar work, Inoue et al. used the same architecture with some modifications to predict the signal strength at 2.1 GHz. In this case, the building occupancy images were developed from aerial maps with the aid of a trained UNet, and the images at the midpoint between the transmitter and receiver were included as the input. The inclusion of midpoint images improved the model's accuracy [33]. However, the model is not suitable for other frequency bands since it was trained with single frequency data.

2.4. Pseudoimages

An improvement to [9] is the formation of a pseudo image composed of the building height shadings in grayscale, feature importance of tabular data, and the satellite image of the receiver location. Each of the three images represented a channel in the pseudo image. The model performed better than when only building height information was used for training and when an XGBoost was trained with the tabular/numerical data [11]. However, the model is not versatile as it is for 900 MHz only, and more detailed information about the path should have been used, not the image at the receiver location alone.

2.5. Summary

Works that use different input, model architectures, as well as outputs were reviewed, and the limitations of the works were also highlighted. One of the limitations is that the idea of predicting a path loss exponent based on any of the image input types will result in erroneous path loss prediction. This is because path loss exponents are estimated based on the least squares method. The process of least squares is associated with its error. By further training a CNN that is not 100% accurate in predicting the path loss exponent, the error during path loss exponent computation is compounded with that of the CNN, thereby affecting the final path loss prediction. Furthermore, using an image representing the receiver environment alone, as was performed in many works, is not sufficient to characterize the propagation path. There are works in which predictions of radio maps were performed based on the UNet architecture. Such predictions are for analysis as images representing the distribution of signal strength/path loss are predicted, not the path loss values. Thus, the outputs are not suitable for automated design and optimization.

On a final note, models valid for multiple frequencies, environments, and antenna heights have not been developed. Radio Frequency Planning Tools are used in wireless network design, which has empirical path loss models available for the designer to choose from. ML path loss models have accuracy better than empirical models, and it is likely that they will replace empirical models in these tools in the future. Hence, it is desirable to have ML models valid for a range of frequencies, antenna heights, and environments. Otherwise, many models have to be developed with varying parameters of frequency, antenna height,

and environment, leading to large memory occupation, time wastage in training individual models, and confusion on the particular model to select from the end user viewpoint. Thus, the study fills the gap of non-availability of versatile models by developing a model for different environments, frequencies, and antenna heights. This will make the ML models flexible for use in different environments, frequencies, and antenna heights. The model considered takes as its input some numeric features affecting path loss combined with features extracted from images by algorithms (CNN or hand-crafted). The image is composed of four satellite images tiled together in a 2 by 2 grid and is for four different positions along the path from transmitter to receiver position. This gives more information on the propagation path than using the image at the receiver location alone. The effects of various feature extraction methods and their combinations on accuracy were investigated and reported. Improvements were observed compared to our previous work in [16]. However, the work considered some datasets in which measurements were taken from networks operating within the UHF band in environments that included rural, suburban, urban, and urban highrise at some fixed values of transmitting and receiving antenna heights.

3. Methodology

3.1. Dataset

The dataset in [16] is used in this work. It is composed of four datasets combined to form a single dataset, with measurements from different frequencies, environments, and antenna heights. Such a type of dataset combination was performed in [11,23]. Three of the datasets are available at [34–36] and have been used in [37–39], respectively. The fourth dataset is measurements taken by us. It should be noted that not all samples in [34] were used to maintain a balanced dataset in terms of environment type. The final dataset is made up of 12369 samples, consisting of 8 features and path loss values as the target. The features include elevation at receiver position (elevation), frequency, height of transmitting antenna (ht), height of receiving antenna (hr), latitude difference between transmitting and receiving antenna (distance_x), longitude difference between transmitting and receiving antenna (distance_y), elevation at transmitter position (tantelev), and distance between transmitter and receiver (distance). Table 1 presents a summary of the dataset, showing the different types of environments from which the measurements were taken and the parameters under which they were taken that include frequency, transmitting antenna height, and receiving antenna height.

Table 1. Dataset summary.

Environment	Frequency (MHz)	Transmitting Antenna Height (m)	Receiving Antenna Height (m)
Rural	868	0.2, 1.5, 3	12
Suburban	1800	30	1.5
Urban	1835.2, 1836, 1840.8, 1864, 2140	40, 41, 53	1, 1.5
Urbanhighrise	868	0.2, 1, 1.5, 3	12

3.2. Image Data

The image used is a concatenation of four satellite images, as illustrated in Figure 1. The images represent various positions along the path from transmitter to receiver location as shown in Figure 1a with a sample in Figure 1b. The images were downloaded from Mapbox [40] based on the latitude and longitude values in the dataset.

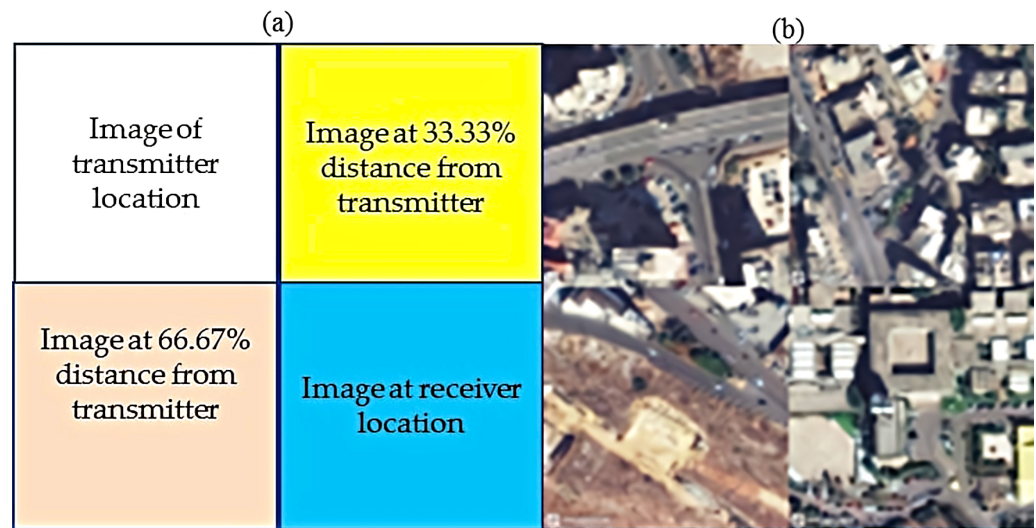


Figure 1. (a) Format of image concatenation. (b) Final concatenated image sample.

3.3. Proposed Method

Figure 2 depicts the development process with three main phases; data preparation, model development and model testing.

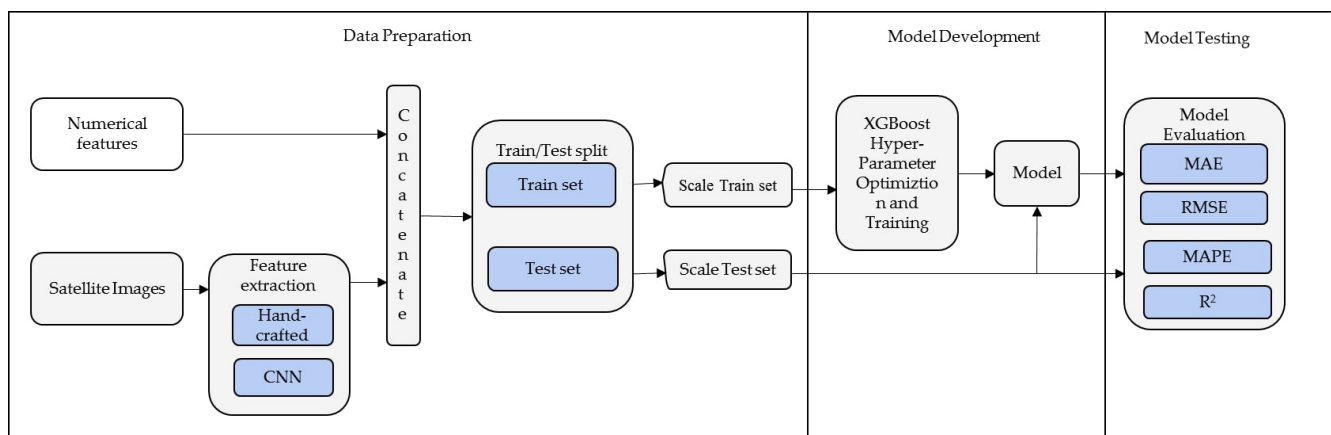


Figure 2. Model development process with three main phases; data preparation, model development, and model testing.

3.3.1. Data Preparation

In the data preparation phase, the features are extracted from satellite images using an algorithm. The extracted features are combined with the numerical features to form a new dataset. The new dataset is then split into the training set and test set with a ratio of 80:20. As a preprocessing step, the two sets of data were scaled with a standard scaler to eliminate magnitude difference amongst features, thereby improving generalization and reducing training time.

3.3.2. Model Development

In the model development phase, hyper-parameter optimization (HPO) was performed on the training set based on XGBoost with hyperband algorithm [41] and then model training was performed with the optimized hyper-parameters.

3.3.3. Model Testing

In this phase, ten different samplings of training and test sets splits were used to train and test an XGBoost fitted with the optimized hyper-parameters. For each of the training

and test sets, model performance based on the test set is noted, and the average of the ten samplings was computed.

3.4. Feature Extraction Methods

The following hand-crafted feature extraction methods were applied to the tiled images: Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG), Gray Level Co-occurrence Matrix (GLCM), and Discrete wavelet transform (DWT). Combinations of GLCM features with other hand-crafted features were considered, as well as a combination of the hand-crafted features with CNN features. The feature extraction methods used are described as follows:

3.4.1. Local Binary Pattern (LBP)

In LBP, a pixel's intensity is compared with other pixel's intensities within its local neighborhood, from which a histogram of binary patterns is obtained [42]. It is a less complex texture extraction method that is invariant to rotation. Differences in pixel intensities are presented using a binary encoding method described in Equation (1).

$$LBP_{(P,R)} = \sum_{p=0}^{p-1} s(g_p - g_c) 2^p \quad (1)$$

where g_c is the center pixel, and g_p is a neighborhood pixel, P stands for the number of neighbors, R is the neighborhood distance, and $s(x)$ is computed in Equation (2) and stands for the binary encoding.

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

LBP features are computed from histograms of the binary encodings based on Equation (3), where $k = 0, 1, 2, \dots, K$, K is the maximum value in the computed $LBP_{(P,R)}$ within the image with dimension (M, N) , and $I(x)$ has a value of 1 when $LBP_{(P,R)}(i, j) = k$, and 0 otherwise. There are various types of LBP [43,44]. A uniform LBP with a p value of 3 and an R -value of 3 was considered in this work.

$$H(k) = \sum_{i=1}^M \sum_{j=1}^N I(LBP_{(P,R)}(i, j)) \quad (3)$$

3.4.2. Gray Level Co-Occurrence Matrix (GLCM)

In GLCM, a matrix $C(i, j)$ known as the co-occurrence matrix is computed by counting the co-occurrences in gray pixels that occur within a distance, d , and orientation, θ , which takes any value between 0 and 315° at 45° intervals. Features can be extracted from the co-occurrence matrix as a form of dimensionality reduction, with Energy, Inertia, Correlation, Difference Moment, Entropy [45], Contrast, and Homogeneity [46], as the most commonly computed features. Many other features can also be extracted from the co-occurrence matrix [47]. A d value of 3 and a θ value of 45° were used, with Contrast, Dissimilarity, Homogeneity, Angular Second Moment (ASM), Energy, and Correlation used as features.

3.4.3. Histogram of Oriented Gradients (HOG)

In HOG, each pixel's magnitude and direction are computed from its x and y gradients. The image is then divided into cells, and histograms are computed for each cell based on gradient values. An aggregation of cells forms a block, and normalization is applied to each block. A combination of the normalized histograms from the blocks forms the HOG features [48]. In this work, the cells are 8 by 8 pixels, and tyeblocks are 2 by 2 cells. The number of orientation bins for the histogram is 9.

3.4.4. Discrete Wavelet Transform (DWT)

In DWT, data are decomposed into low pass L and high pass H components, yielding two sub-bands of $L/H = l_1, l_2, \dots, l_{(N/2)}, h_1, h_2, \dots, h_2$. For a bidimensional image, decomposition is performed in the vertical and horizontal directions, resulting in four sub-bands. The formed sub-bands are Approximation, Horizontal, Vertical, and Diagonal details and are due to the combination of LL , LH , HL , and HH filters, respectively. Statistical parameters can be computed from each sub-band to have a reduced feature set [43,47]. Methods such as the Line Measure Technique (LMT) [49] are also used to reduce the feature set. The mean, energy, standard deviation, and entropy of each sub-band were used as the extracted features in this work, summing to a total of 16 features.

3.4.5. Segmentation Fractal-Based Texture Analysis (SFTA)

SFTA is a two-staged feature extraction technique where, in the first stage, Two-Threshold Binary Decomposition (TTBD) is applied on a grayscale image to produce binary images. The Multilevel Otsu algorithm is used in TTBD to calculate some threshold values from the image's gray level distribution. The algorithm first determines a threshold value that ensures that the image's intra-class variance is minimized. This is applied to all regions of the image to obtain n_t threshold values. Decomposition is carried out by the selection of pairs of threshold values and the application of a two-threshold segmentation, as shown in Equation (4) [50].

$$I_b(x, y) = \begin{cases} 1 & t_l < I(x, y) \leq t_u \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$I(x, y)$ and $I_b(x, y)$ are pixel intensities of grayscale and binary images, respectively, at (x, y) pixel position. t_l and t_u are the upper and lower threshold values. A total of $(n_t - 1)$ binary images are produced based on the procedure. In addition, simple thresholding is applied to produce n_t binary images summing up to $(2n_t - 1)$ binary images. In the second step, three features are extracted from each binary image, namely, Area (A), Mean Gray Level (V), and Fractal Dimension (D). Area (A) represents the number of white pixels in the binary image, Mean Gray Level (V) is the average brightness of gray image pixel positions corresponding to white pixel positions in the binary image, and Fractal Dimension (D) is a measure of boundary complexity. A total of $(6n_t - 3)$ features are extracted for an input image in SFTA [18]. An n_t value of 8 was used, thereby extracting 45 features from each input image.

3.4.6. Convolutional Neural Networks (CNN)

A CNN is used to process data that is in grid form, such as an image. It is made up of three main layers, the convolutional layer, the pooling layer, and the fully connected layer. Feature extraction is performed at the convolutional and pooling layers, whereas the fully connected layer performs mappings between the extracted features and target values. The convolutional layer is made up of low-dimensional filters that are multiplied by the image pixels by sliding it from left to right and top to bottom based on a stride value. A stride is the number of skips in pixel values for the filter to slide across the image. Each convolutional layer is followed by activation, and the Rectified Linear Unit (ReLU) is commonly used because it is fast and gives a better generalization. The image dimension is reduced when passed through a convolutional layer. The pooling layer is for further dimension reduction to decrease training time and increase immunity against noise, thereby preventing overfitting. Pooling could be carried out by summing the pixel values within a small grid (Sum Pooling), averaging (Average Pooling), or finding the maximum value (Max Pooling). Convolution layers and pooling layers are stacked for feature extraction. The deeper the network, the more sophisticated the process of extracting features. The first convolutional layer extracts features such as lines and edges, and deeper layers extract more detailed information. The fully connected layer is made up of neurons, as in the case of a Multiple Layer Perceptron, where two or more layers could be stacked, with the last

layer forming the output layer. Regularization techniques, such as image augmentation, dropout, and batch normalization, are used to improve the accuracy of CNN [51–53]. The CNN used in this work is made up of five convolutional layers with batch normalization, and maximum pooling applied after each layer. The output of the last max pooling layer is passed to a fully connected layer with 177 neurons from which the features are extracted. The number of filters in the five consecutive convolutional layers are 24, 38, 83, 98, and 199, respectively, with square filters of sizes 2, 4, 4, 4, and 3, respectively. The maximum pooling layers used are of size 2. The complexity of the CNN is less compared to those of pretrained networks such as VGG-16, ResNet50V2, InceptionV3, EfficientNetB0, and so on. It, therefore, trains faster and occupies less memory space.

3.5. Extreme Gradient Boosting (XGBoost)

XGBoost is an advanced Gradient Boosting with improvements in speed, efficiency, and scalability. It has produced high accuracy in various applications such as prediction of Type 2 diabetes risk [54], hot spots at protein–protein interfaces [55], sales forecasting [56], and has been used in solving various problems in Kaggle competitions [57]. Like any boosting algorithm, it is made up of Classification and Regression Trees (CART) as base learners. Sometimes, linear functions are also used as base learners. The base learners are fitted to minimize a loss function. The loss function determines how well the model performs, and it is minimized during training [58,59]. Minimization of the loss function is achieved through gradient descent [60]. In XGBoost, a regularization term is added to the loss function, and Taylor Series Expansion is used in approximating the modified loss function. Optimization of the tree structure is performed using an approximate greedy algorithm. XGBoost uses column subsampling such that portions of features are collected to grow trees, thereby minimizing overfitting and improving training speed. In addition, it uses a presorting technique and storing mechanism whereby datasets are stored in blocks, and the search for the best split is performed by scanning columns in a block. It also uses techniques such as Block Compression and Block Sharding for out-of-core computations and when fetching data from multiple drives, respectively, thereby enhancing the execution time [58,61]. The boosting system can be composed of normal trees, as in the case of Gradient Boosting, or the use of Dropouts meet Multiple Additive Regression Trees (DART). DART is a process whereby multiple numbers of regression trees are fitted at a low learning rate, and trees fitted earlier are maintained, whereas those fitted later are dropped in order to prevent overfitting [62]. In this work, the number of trees, learning rate, and type of booster were ascertained based on hyperband for each of the combinations of numeric features with one or two feature extraction methods used on the input image, as shown in Table 2.

Table 2. Hyper-parameters of XGBoost for combinations of numeric features and features from one or two feature extraction method(s).

Feature Extraction Method	Number of Trees	Learning Rate	Booster
No Image	248	0.2912	gbtree
LBP	209	0.2698	dart
HOG	260	0.1035	dart
GLCM	251	0.2644	gbtree
DWT	261	0.1375	gbtree
SFTA	202	0.1582	gbtree
GLCM + LBP	263	0.1649	dart
GLCM + HOG	287	0.1178	dart
GLCM + DWT	276	0.2807	dart
GLCM + SFTA	219	0.1994	dart
CNN	100	0.3000	gbtree
CNN + LBP	195	0.1704	dart
CNN + HOG	185	0.0436	dart
CNN + GLCM	293	0.0908	gbtree
CNN + DWT	209	0.0711	dart
CNN + SFTA	199	0.0997	gbtree

3.6. Performance Metrics

The performance metrics used in evaluating the performance of models include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2) [63,64].

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_{m(i)} - P_{p(i)}| \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(P_{m(i)} - P_{p(i)})^2}{N}} \quad (6)$$

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{P_{m(i)} - P_{p(i)}}{P_{m(i)}} \right| \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (P_{m(i)} - P_{p(i)})^2}{\sum_{i=1}^N (P_{m(i)} - \bar{P}_p)^2} \quad (8)$$

P_m is the measured path loss, P_p is the predicted path loss, $\bar{P}_p$ is the mean of predicted path loss values, and N is the number of samples.

4. Result and Discussions

The performance metrics of each model resulting from a combination of non-image features merged with features from one or two feature extraction techniques applied to the images are shown in Table 3. Amongst the hand-crafted features, GLCM had the lowest error metrics and the highest coefficient of determination, but its accuracy is below the case in which no image features were used. The combination of GLCM features with other hand-crafted features further decreased the accuracy below that of GLCM features. The addition of hand-crafted features led to a reduction in accuracy when compared to using only non-image features, as shown by the performance metrics as well as the last column of Table 3, indicating a percentage improvement over using only non-image features. Negative values were observed as percentage improvements, indicating a reduction in performance. However, even with the reduction in performance in models with hand-crafted features, they are expected to perform better when tested with data from a site not used in model training. This is because the model trained with the hand-crafted features has some knowledge of the environments. The performance when CNN-extracted features were used was better than when features extracted based on any of the hand-crafted techniques were used. A combination of CNN features with any hand-crafted feature extraction methods decreased the accuracy compared to when CNN features were used alone, except in the case of LBP and GLCM, where an increase was observed. A combination of CNN-extracted features with GLCM features (CNN + GLCM) had the highest overall accuracy with an R^2 of 0.9372, seconded by a combination of CNN with LBP (CNN + LBP) with an R^2 of 0.9364 and were ahead of CNN features ($R^2 = 0.9357$). The RMSE value of 3.6682 dB when CNN + GLCM features were used is an indication that the model's performance is acceptable as it is below the maximum allowed value. The acceptable RMSE for a path loss model is a value less than 7 dB in an urban environment and any value less than 15 dB in suburban and rural environments [65]. The RMSE value achieved is an improvement over the results in [11] with an RMSE of 8.27 dB, [17], 6.8944 dB, [12], 4.4286 dB, [29], 7.61 dB, and [66], 3.89 dB.

The results indicated that the hand-crafted features and their combinations could not compete with CNN for feature extraction. As such, the requirement of training a CNN that needs a large volume of data and training time still holds as CNN-extracted features cannot be replaced by hand-crafted features due to the low accuracy of the hand-crafted features. However, the addition of LBP or GLCM features to the CNN features improved model accuracy over CNN features alone. However, since a combination of CNN

and GLCM features had the highest improvement; they are recommended for use in this application. Meanwhile, the architecture of the CNN used resulted in an 18-fold speed over the ResNet50V2 used in [16] as it trained in 29.8242 min while the ResNet50V2 trained in 546.5685 min. In addition, a reduction in the RMSE value of 0.7604 dB was observed compared to the results in [16]. Compared to models based on diffraction that require the complete path profile from transmitter to receiver, such as the Walfisch Ikegami and Walfisch Bertoni, the proposed model is cost-effective and saves time as it does not require the additional measurement of parameters such as building heights and street width. In addition, ML models have accuracies better than such models, as seen in [6,67,68] because ML models do capture non-linear relationships.

Table 3. XGBoost performance on combinations of numeric features and features from one or two feature extraction method(s).

Feature Extraction Method	MAE (dB)	RMSE (dB)	MAPE (%)	R ²	% Improvement
No Image	2.9862	4.0500	2.2985	0.9227	NA
LBP	3.1914	4.2985	2.4589	0.9137	−6.1358
HOG	3.4590	4.7434	2.6751	0.8949	−17.1210
GLCM	3.0964	4.1616	2.3852	0.9191	−2.7557
DWT	3.2247	4.3281	2.4897	0.9125	−6.8667
SFTA	3.2831	4.4040	2.5312	0.9094	−8.7407
GLCM + LBP	3.2035	4.3003	2.4709	0.9136	−6.1803
GLCM + HOG	3.4347	4.7172	2.6530	0.8961	−16.4741
GLCM + DWT	3.2049	4.3393	2.4688	0.9120	−7.1432
GLCM + SFTA	3.2309	4.3358	2.4880	0.9123	−7.0568
CNN	2.6268	3.7106	2.0324	0.9357	8.3802
CNN + LBP	2.6129	3.6909	2.0208	0.9364	8.8667
CNN + HOG	2.7698	4.1205	2.1484	0.9207	−1.7407
CNN + GLCM	2.5848	3.6682	2.0013	0.9372	9.4272
CNN + DWT	2.6305	3.7882	2.0394	0.9329	6.4642
CNN + SFTA	2.6303	3.7655	2.0368	0.9338	7.0247

Deterministic models, which also use the complete profile from transmitter to receiver, as well as the principles of wave transmission, such as reflection, refraction, diffraction, and shadowing, have high computational complexity and are very slow. With the proposed method, predictions are made instantly from the trained model, and the model can be used for similar environments. Meanwhile, in the case of deterministic models, each environment has to undergo the process of computation.

Figure 3 shows the percentage improvement calculated from RMSE values when either CNN, CNN + LBP, or CNN + GLCM features were used in combination with numeric features over when numeric features were used alone. The addition of CNN + GLCM features to the numeric features had the highest improvement of 9.4272%, followed by the addition of CNN + LBP, which resulted in 8.8667%, and lastly, the addition of CNN features improved it by 8.3802%, as shown in line Table 3.

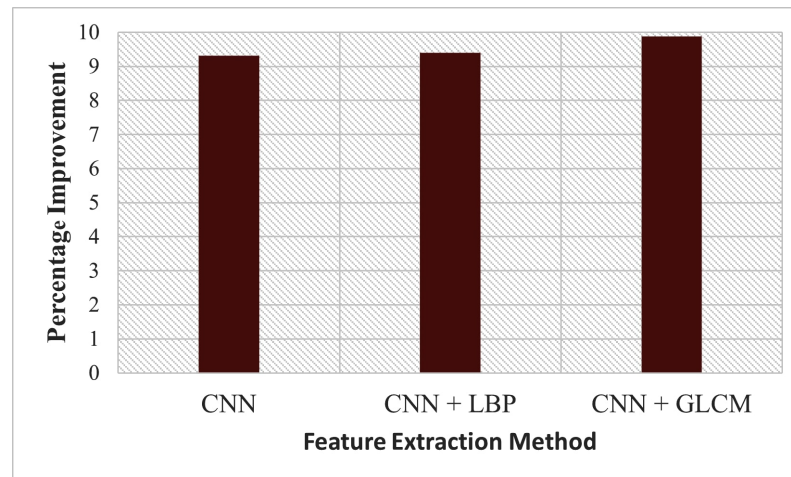


Figure 3. Percentage improvement over non use of images for the top three feature extraction methods; CNN, CNN + LBP, and CNN + GLCM.

4.1. Comparison of XGBoost with Other Algorithms

The best features observed were a combination of the numeric features and features extracted using GLCM and CNN. Four other models were developed based on the features extracted using different algorithms, and the performances of the models were compared with XGBoost. The algorithms considered are Random Forest (RF), Extreme Learning Trees (ET), Gradient Boosting (GB), and K Nearest Neighbor (KNN), with hyper-parameter optimization conducted in each case. The average RMSE of each algorithm is presented in Figure 4. It is observed that XGBoost (together with CNN and GLCM extracted features and numeric features) had the lowest RMSE, followed by GB, and ET had the highest value. XGBoost was better because of the regularization term added to its loss function and the column subsampling technique employed in it, which minimizes overfitting. ET had the least performance because it does not employ any optimization process for the determination of splits; rather it chooses the splits randomly.

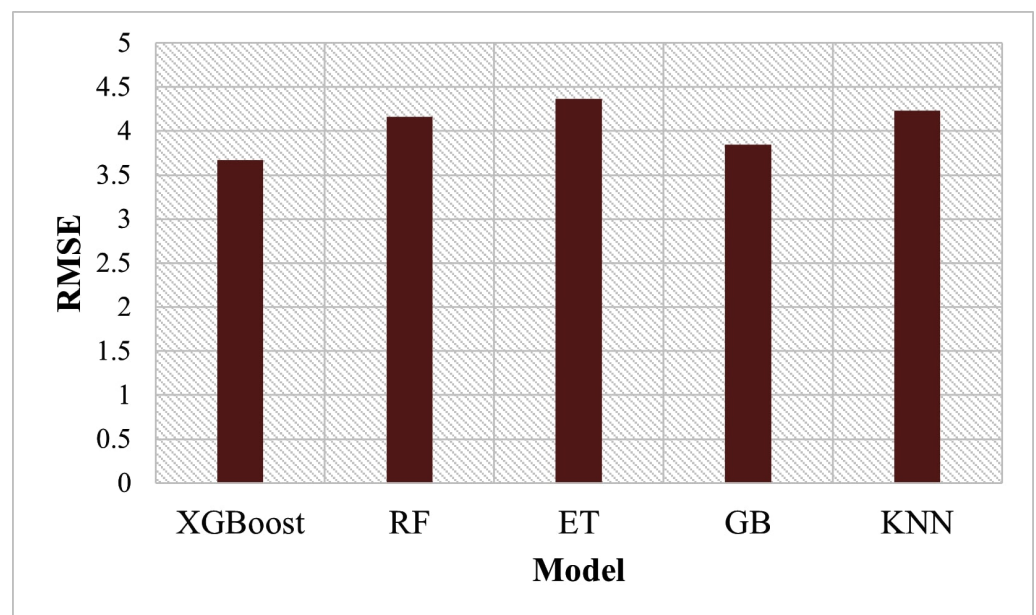


Figure 4. RMSE values of the ML models for a combination of CNN + GLCM features and non-image features.

4.2. Feature Importance

Feature importance of the best model (with CNN + GLCM features), as well as the model that has only non-image features, were determined based on the Shapley Additive Explanations (SHAP). SHAP is a method for interpreting ML models, based on a combination of game theory, Shapley values, and some conditional expectations to compute the importance of each feature [69]. Figures 5–8 present explanations of the model with non-image features only. Figure 5 shows the mean importance of each feature, with “tannaelev” having the highest importance. Next to it is “distance”, as expected, because it is well established that path loss varies with distance. However, “tannaelev” had the highest impact on the model output because studies have shown that apart from distance, the elevation of a position affects path loss at that position due to variations in height difference between the transmitting and receiving antennas. Thus, for a multiple antenna path loss model, it is wise to include the elevation of the transmitting and receiving antenna positions so that their height difference effect is captured. The “tannaelev” had an importance that superseded that of “distance” because it was observed by Jimoh et al. [70] that elevation affects path loss to the extent that a position at a further distance from the transmitting antenna but at a high elevation could have lower path loss values than a position closer to the transmitting antenna but at a low elevation. The third and fourth most important features are distances in latitude and longitude, and then the height of the receiving antenna (hr) had the least importance.

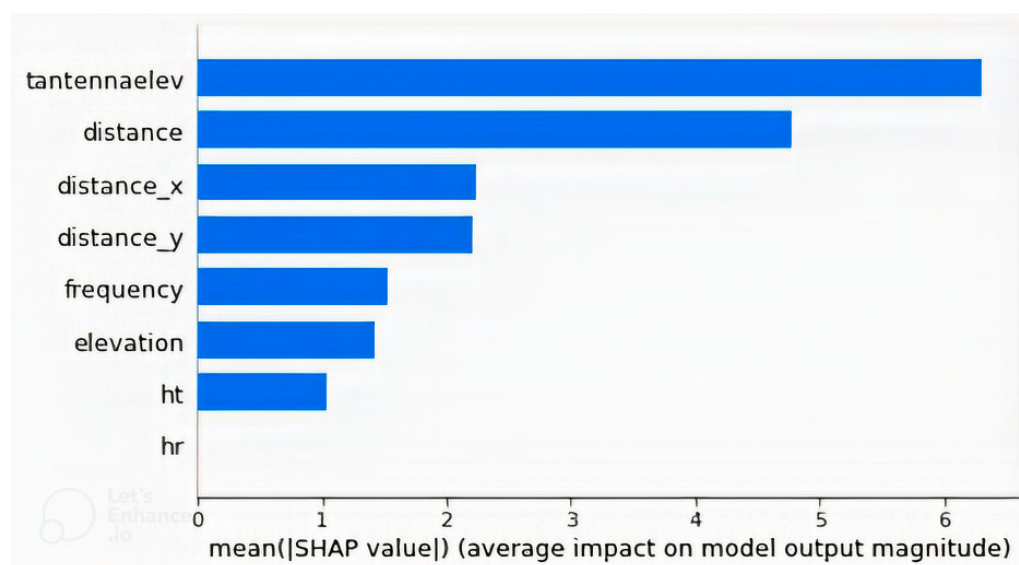


Figure 5. Mean feature importances of the model with non-image features only, with “tannaelev” having the highest importance and “hr” the least.

Figure 6 shows the detailed explanations of the features, with the position of the dots with respect to the horizontal axis showing SHAP values that determine whether the path loss is high or low, and the color indicating the feature value. Positive SHAP values mean that the feature has a positive impact on path loss and negative means it has a negative impact. The results show that for higher values of “tannaelev”, the path loss value is low, but for lower feature values, the output is high, though with a few instances where the output is low at lower values of “tannaelev”. This is because the higher the elevation of the transmitting antenna, the wider the signal coverage; therefore, path loss decreases for the distance concerned. Likewise, the lower the elevation, the smaller the coverage, and the higher the path loss values. The fewer points where the low output values were observed might be due to irregularities in elevations of the receiver position as well as clutter. In the case of “distance”, it is observed that at lower distances, the path loss values are low and high at higher distances. Some minor instances in which the linearity is violated are due to the effect of other features. However, not only does the difference between transmitting and

receiving antenna altitude (height + elevation) affect the signal coverage. Other antenna parameters such as azimuth and tilt also affect the signal coverage, but the study of their effects is beyond the scope of this work.

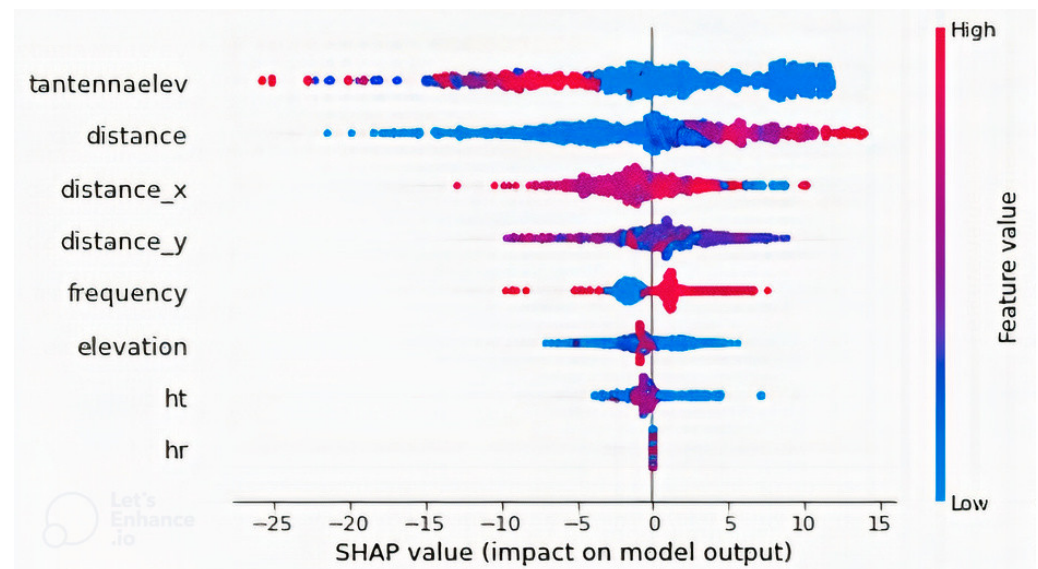


Figure 6. SHAP summary plot of the model with non-image features only, showing the variations in feature values with output path loss value.

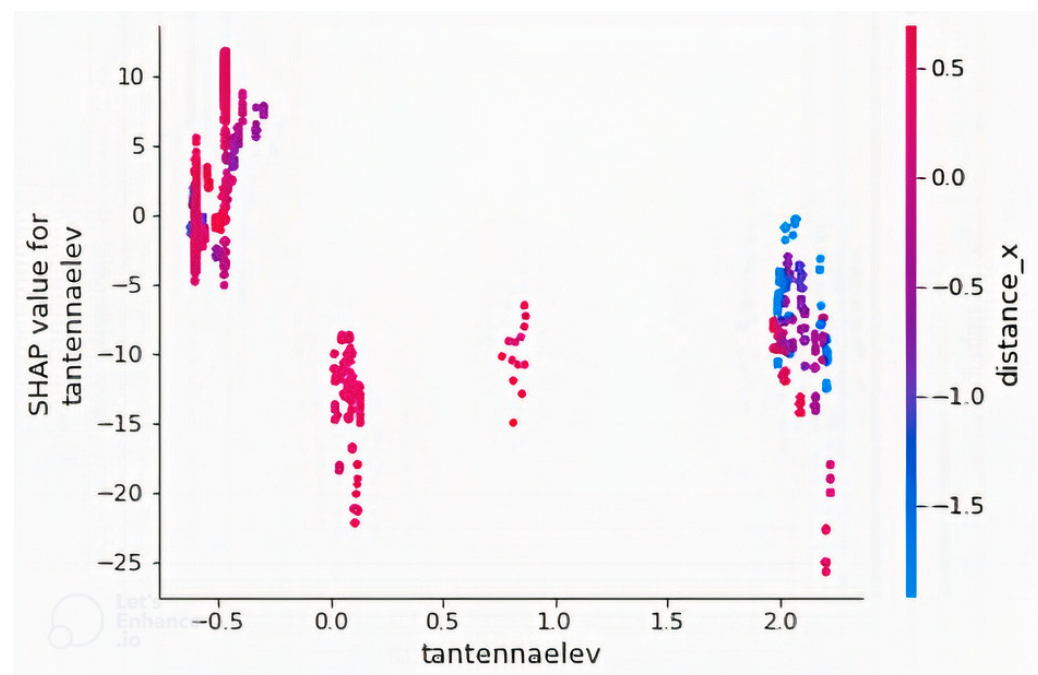


Figure 7. Dependence plot of the feature “tantennaelev” in a model with non-image features only with the feature “distance_x”. Feature “tantennaelev” has the highest interaction index with “distance_x”.

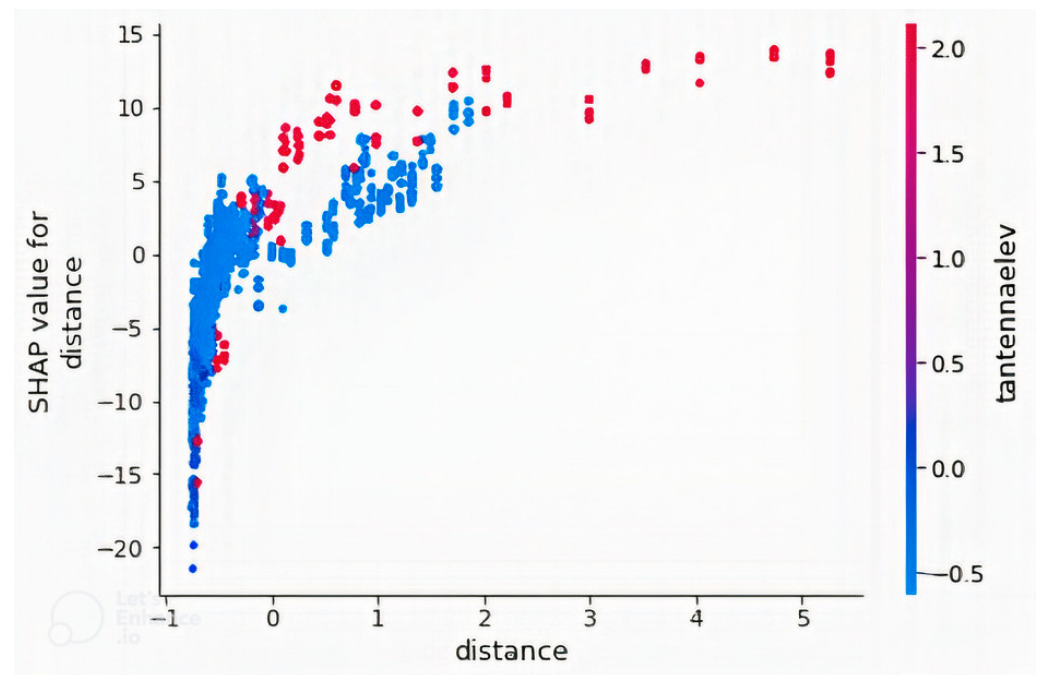


Figure 8. Dependence plot of the feature “distance” in a model with non-image features only with feature “tannaelev”. Feature “distance” has the highest interaction index with “tannaelev”.

The difference in latitude between the transmitting and receiving antenna and that of longitude (distance_x and distance_y) were observed to push path loss values high and moderate, respectively. Higher “frequency” values pushed the path loss values higher as expected, the moderately low values pushed it low, and a few instances in the lowest frequency values pushed the path loss high due to the effect of other factors, such as elevation and/or clutter. The features “elevation” and height of the receiving antenna (ht) were observed to push the path loss values low when they are high or extremely low, whereas the upper part of their low values increased path loss values. This could be due to situations when the transmitting antenna is at a higher elevation.

Dependence plots of the two most important features (tannaelev and distance) were plotted. Figure 7 shows that “tannaelev” has some interactions with “distance_x”. Therefore, the effect of “tannaelev” on path loss is dependent on “distance_x” value. The realization that “distance_x” has interaction with “tannaelev” is based on the computation of the approximate interaction index between the features, from which the value between “tannaelev” and “distance_x” was the highest. Confirmation of the presence of interaction is based on the presence of some vertical color patterns. Conclusions about the interactions of these variables cannot be made based on the plot because the data used are not a representation of all latitude and longitude positions on earth. Figure 8 is the dependence plot of “distance”, showing how path loss increases with increasing distance. The presence of vertical color patterns within the plot around the start values of distance indicates an interaction between the features “distance” and “tannaelev”. As mentioned earlier, path loss at a distance is dependent on elevation. Therefore, path loss at the same distance but different elevation values are unequal.

Figure 9 shows the 20 topmost features in terms of mean importance in the model that had non-image inputs as well as combinations of features extracted using CNN and GLCM from the images. Features named “f_x” are the features extracted by the CNN, where x takes values from 0 to 176. It is observed that the 144th feature “f_143” had the highest importance, followed by the third, “f_2”. Only two of the non-image features appeared in the top 20, with the height of the transmitting antenna (ht) appearing as the 13th and elevation as the 20th. The most important features are the image features extracted using CNN, and none of the GLCM features appeared in the top 20. Unlike in the case of the non-

image input model, where “ht” and “elevation” were the last two, in this case, they were the first two because, in the non-image input model, the model was developed without consideration of environmental factors, and its explanation was similar to that of existing empirical models. The inclusion of features of the satellite images enhances the model’s ability to differentiate between environments. The model is well fitted to differentiate between the environments based on the importance it gives to the extracted features.

Figure 10 is a summary plot of the 20 most important features. It was observed that higher values of the CNN features produce higher path loss values, and lower values produce lower path loss values, with the exceptions of “f_0” and “f_1”. However, the features were extracted by a CNN, and the reason for this behavior cannot be explained as the features are not physical quantities and were not generated by a known equation but a black box model, unlike other quantities such as “distance” and “elevation” that are physical quantities. However, it is clearly seen that predicted path loss is dependent on these extracted features. Similarly, dependence plots for features “f_143” and “f_2” are presented in Figures 11 and 12, respectively. Figure 11 showed that “f_143” had the highest approximate interaction index with “f_167”, and it was observed that as “f_143” increased, “f_167” also increased, as well as path loss but in a stepwise fashion without interactions even though their partial interaction index is the highest. The pair of feature “f_2” as shown in Figure 12 is “correlation”, which is a GLCM feature. Interactions were found between the variables due to the presence of vertical color patterns, and different portions of the plots show different combination rules that determine path loss values. The vertical color patterns indicate that for the same “f_2” value, the path loss value differs depending on the value of “correlation”. More of such relationships between features could be observed if the dependence plot of the remaining features is plotted. The concept of feature interaction can be explained by Equation (9), where y depends on two variables x_1 and x_2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \text{error} \quad (9)$$

β_0 is an average response, β_1 and β_2 define contributions of features x_1 and x_2 , respectively, and β_3 defines the combined contributions of x_1 and x_2 beyond their individual explanations alone. Random variations in the values unexplained by the remaining part of the equation are represented by the error term. Thus, the interaction between two features such as “f_2” and “correlation” can be combined and included in an explainable model similar to the $\beta_3 x_1 x_2$ term in Equation (9) by replacing x_1 and x_2 with “f_2” and “correlation”, and the dependent variable y with path loss. Further analysis of interactions between the features could result in a glass box path loss model in a form similar to Equation (9) that gives an explanation of how the path loss values are calculated, unlike ML models that are black box models [71,72].

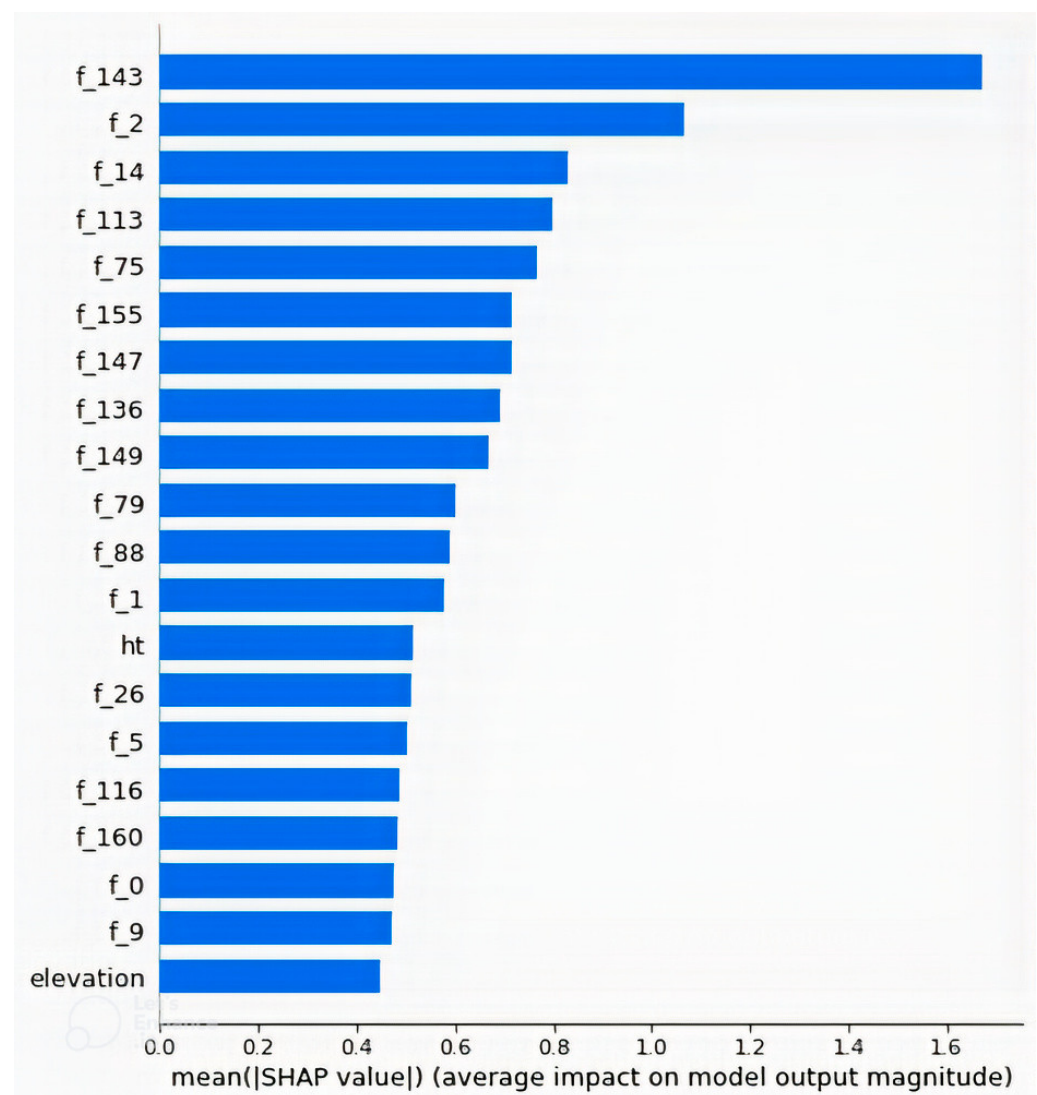


Figure 9. Mean feature importances of the top 20 features in the model with a combination of non-image features and CNN + GLCM features. “f_143”, which is a CNN-extracted feature, has the highest importance, while “elevation”, which stands for elevation at receiving antenna position, has the least importance among them.

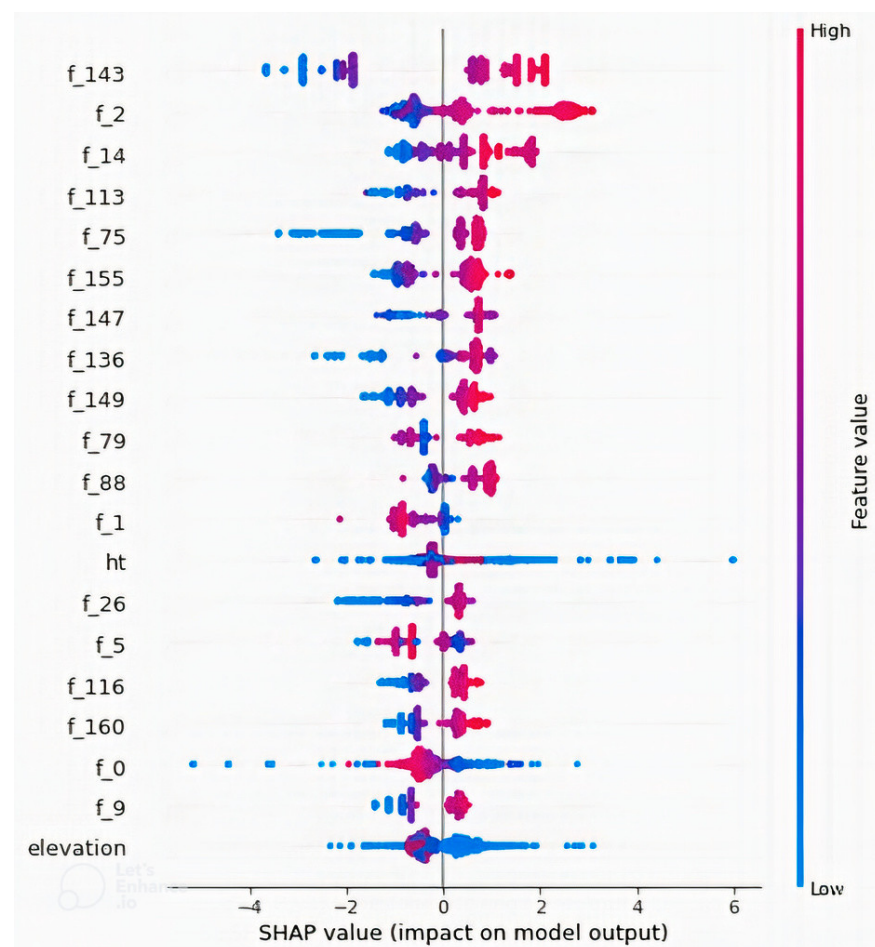


Figure 10. SHAP summary plot of the 20 most important features in the model developed with a combination of non-image features and CNN + GLCM features. It shows the variations in feature values with output path loss value.

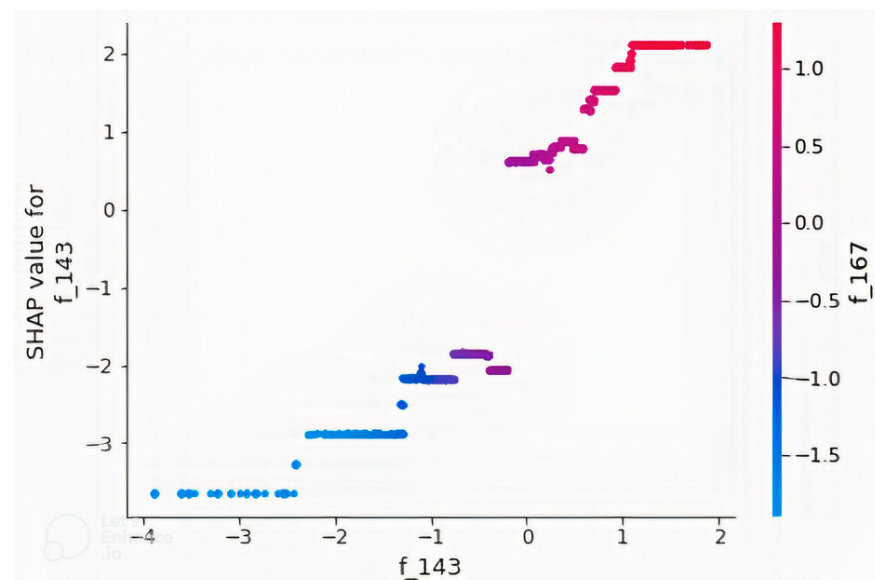


Figure 11. Dependence plot of the feature “f_143” in the model developed with a combination of non-image features and CNN + GLCM features. Feature “f_143” has the highest interaction index with “f_167”.

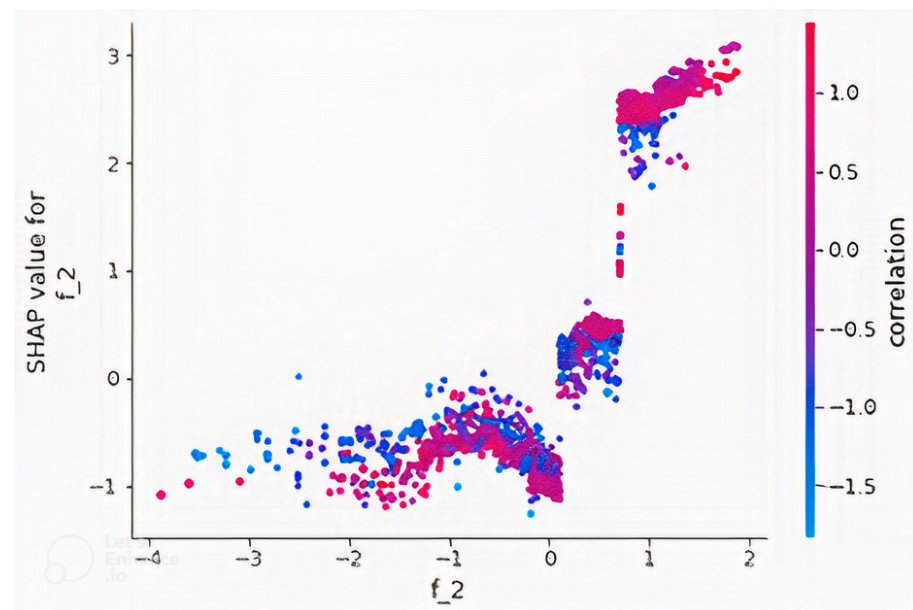


Figure 12. Dependence plot of the feature “f_2” in the model developed with a combination of non-image features and CNN + GLCM features. Feature “f_2” has the highest interaction index with the GLCM feature “correlation”.

5. Conclusions

A path loss model with a good prediction accuracy is required for the determination of wireless network parameters such as transmitting power, cell radius, antenna height, number of cell sites, and their optimal position. The attainment of the required data rate is dependent on the proper setting of the parameters such that there is good signal coverage and no interference. The dependency of path loss on the environment led to the inclusion of satellite images of network location in developing machine learning path loss models such that the influence of the environment is learned. A path loss predictive model that handles multiple frequencies, multiple antenna heights, and multiple environments were developed in this work. The model takes in a combination of numeric features and features extracted from satellite images to make the path loss prediction. XGBoost was the algorithm used in model development and was selected based on its good performance in several applications. Several hand-crafted feature extraction methods, their combinations, or combinations of hand-crafted features and CNN features, were tested. The hand-crafted feature extraction methods included LBP, HOG, GLCM, DWT, and SFTA. None of the hand-crafted features improved accuracy compared with using numeric features alone; instead, the accuracy was reduced. Combining GLCM with any of the hand-crafted feature extraction methods decreased performance below that of GLCM features. When CNN features were combined with any of the hand-crafted feature extraction methods, combinations with LBP or GLCM resulted in accuracies better than that of CNN features. A combination of CNN and GLCM features from the images had the best performance overall. This resulted in a 9.4272% improvement in RMSE for combined environments over when no satellite image features were used in model development. A comparison was performed to evaluate the performance of other ML algorithms on the combination of numeric, GLCM, and CNN features. Random Forest, Extreme Learning Trees, Gradient Boosting, and K Nearest Neighbor were used for this purpose. XGBoost had the best performance metrics and also has an advantage of a fast training time over them. Features extracted using CNN were observed to have the highest impact on the model’s output based on SHAP. Although the model’s accuracy is good, there is a need for more training data from different environments and network parameters such that a single model can be developed and used for pre-network deployment path loss prediction. Meanwhile, the study demonstrated the possibility of developing such a model because an RMSE value of

3.6682 dB was obtained when combined data were used. This RMSE value is acceptable for a path loss model as it is below 7 dB. In our future work, the model's performance for the different sub-datasets will be explored, and methods for improving the accuracy of sub-datasets with low accuracy will be addressed.

Author Contributions: Conceptualization, U.S.S., D.T.C.L. and O.A.M.; methodology, U.S.S., D.T.C.L. and O.A.M.; formal analysis, U.S.S.; data curation, U.S.S.; writing—original draft preparation, U.S.S.; writing—review and editing, U.S.S., D.T.C.L. and O.A.M.; supervision, D.T.C.L. and O.A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Universiti Brunei Darussalam, grant number RSCH/1.18/FICBF(a)/2022/004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this research are available on the github repository, "A Multiple Environment and Multiple Network Parameter Path Loss Dataset".

Acknowledgments: We express our sincere gratitude to Universiti Brunei Darussalam for the provision of resources used in this research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lukman, S.; Nazaruddin, Y.Y.; Ai, B.; He, R.; Joelianto, E. Estimation of received signal power for 5G-railway communication systems. In Proceedings of the 6th International Conference on Electric Vehicular Technology (ICEVT 2019), Bali, Indonesia, 18–21 November 2019; pp. 35–39.
2. Jo, H.S.; Park, C.; Lee, E.; Choi, H.K.; Park, J. Path Loss Prediction Based on Machine Learning Techniques: Principal Component Analysis, Artificial Neural Network and Gaussian Process. *Sensors* **2020**, *20*, 1927. [\[CrossRef\]](#)
3. Zhang, Y.; Wen, J.; Yang, G.; He, Z.; Wang, J. Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion. *Appl. Sci.* **2019**, *9*, 1908. [\[CrossRef\]](#)
4. Sotiroudis, S.P.; Goudos, S.K.; Siakavara, K.F.I.A. Tool to Explain Radio Propagation and Reduce Model Complexity. *Telecom* **2020**, *1*, 114–125. [\[CrossRef\]](#)
5. Moraitis, N.; Tsipi, L.; Vouyioukas, D.; Gkioni, A.; Louvros, S. Performance Evaluation of Machine Learning Methods for Path Loss Prediction in Rural Environment at 3.7GHz. *Wirel. Netw.* **2021**, *27*, 4169–4188. [\[CrossRef\]](#)
6. Isabona, J.; Imoize, A.; Ojo, S.; Karunwi, O.; Kim, Y.; Lee, C.; Li, C. Development of a Multilayer Perceptron Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments. *Appl. Sci.* **2022**, *12*, 5713. [\[CrossRef\]](#)
7. Ahmadien, O.; Ates, H.F.; Baykas, T.; Gunturk, B.K. Predicting Path Loss Distribution of an Area from Satellite Images Using Deep Learning. *IEEE Access* **2020**, *8*, 64982–64991. [\[CrossRef\]](#)
8. Thrane, J.; Zibar, D.; Christiansen, H.L. Model-Aided Deep Learning Method for Path Loss Prediction in Mobile Communication Systems at 2.6 GHz. *IEEE Access* **2020**, *8*, 7925–7936. [\[CrossRef\]](#)
9. Sotiroudis, S.P.; Goudos, S.K.; Siakavara, K. Deep Learning for Radio Propagation: Using Image-Driven Regression to Estimate Path Loss in Urban Areas. *ICT Express* **2020**, *6*, 160–165. [\[CrossRef\]](#)
10. Lee, J.G.Y.; Kang, M.Y.; Kim, S.C. Path loss exponent prediction for outdoor millimeter wave channels through deep learning. In Proceedings of the IEEE Conference on Wireless Communications and Networking, Marrakech, Morocco, 15–18 April 2019.
11. Sotiroudis, S.P.; Sarigiannidis, P.; Goudos, S.K.; Siakavara, K. Fusing Diverse Input Modalities for Path Loss Prediction: A Deep Learning Approach. *IEEE Access* **2021**, *9*, 30441–30451. [\[CrossRef\]](#)
12. Omoze, E.L.; Edeko, F.O. Statistical Tuning of COST 231 Hata Model in Deployed 1800 MHz GSM Networks for a Rural Environment. *Niger. J. Technol.* **2021**, *39*, 1216–1222. [\[CrossRef\]](#)
13. Cahyadi, M.B.; Sudiarta, P.K.; Hartawan, D.D.; Analisis, I.G.A.K. Perbandingan Nilai Shadow Fading Pada Model Propagasi Stanford University Interim (Sui) Dengan Metode Simulasi Dan Drive Test. *J. Spektrum* **2021**, *8*, 230–242.
14. Ayadi, M.; Zineb, B.; Tabbane, A.; Uhf, S.A. Path Loss Model Using Learning Machine for Heterogeneous Networks. *IEEE Trans. Antennas Propag.* **2017**, *65*, 3675–3683. [\[CrossRef\]](#)
15. Nguyen, C.; Cheema, A.A. A Deep Neural Network-Based Multi-Frequency Path Loss Prediction Model from 0.8 GHz to 70 GHz. *Sensors* **2021**, *21*, 5100. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Sani, U.S.; Lai, D.T.C.; Malik, O.A. A hybrid combination of a convolutional neural network with a regression model for path loss prediction using tiles of 2D satellite images. In Proceedings of the 8th International Conference on Intelligent and Advanced Systems (ICIAS), Kuching, Malaysia, 13–15 July 2021.

17. Sani, U.S.; Lai, D.T.C.; Malik, O.A. Investigating Automated Hyper-Parameter Optimization for a Generalized Path Loss Model. In Proceedings of the 11th International Conference on Electronics Communications and Networks (CECNet), Beijing, China, 18–21 November 2021; pp. 283–291.
18. Sotiroidis, S.P.; Siakavara, K.; Koudouridis, G.P.; Sarigiannidis, P.; Goudos, S.K. Enhancing Machine Learning Models for Path Loss Prediction Using Image Texture Techniques. *IEEE Antennas Wirel. Propag. Lett.* **2021**, *20*, 1443–1447. [\[CrossRef\]](#)
19. Bodapati, J.D.; Suvarna, B.; Veeranjanyulu, N. Role of Deep Neural Features vs Hand Crafted Features for Hand Written Digit Recognition. *Int. J. Recent Technol. Eng* **2019**, *7*, 147–152.
20. Sumi, T.A.; Hossain, M.S.; Islam, R.U.; Andersson, K. Human Gender Detection from Facial Images Using Convolution Neural Network. In Proceedings of the Communications in Computer and Information Science, Nottingham, UK, 30–31 July 2021; Volume 1435, pp. 188–203.
21. Lin, W.; Hasenstab, K.; Cunha, G.M.; Schwartzman, A. Comparison of Handcrafted Features and Convolutional Neural Networks for Liver MR Image Adequacy Assessment. *Sci. Rep.* **2020**, *10*, 20336. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Nanni, L.; Ghidoni, S.; Brahnam, S. Handcrafted vs. Non-Handcrafted Features for Computer Vision Classification. *Pattern Recognit.* **2017**, *71*, 158–172. [\[CrossRef\]](#)
23. Thrane, J.; Sliwa, B.; Wietfeld, C.; Christiansen, H.L. Deep learning-based signal strength prediction using geographical images and expert knowledge. In Proceedings of the 2020 IEEE Global Communications Conference, Taipei, Taiwan, 7–11 December 2020.
24. Sliwa, B.; Geis, M.; Bektas, C.; Lop, M.; Mogensen, P.; Wietfeld, C. DRaGon: Mining latent radio channel information from geographical data leveraging deep learning. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022.
25. Nguyen, T.T.; Caromi, R.; Kallas, K.; Souryal, M.R. Deep learning for path loss prediction in the 3.5 GHz CBRS spectrum band. In Proceedings of the 2022 IEEE Wireless Communications and Networking Conference (WCNC), Austin, TX, USA, 10–13 April 2022.
26. Ates, H.F.; Hashir, S.M.; Baykas, T.; Gunturk, B.K. Path Loss Exponent and Shadowing Factor Prediction From Satellite Images Using Deep Learning. *IEEE Access* **2019**, *7*, 101366–101375. [\[CrossRef\]](#)
27. Cheng, H.; Lee, H.; Cho, M. Millimeter Wave Path Loss Modeling for 5G Communications Using Deep Learning With Dilated Convolution and Attention. *IEEE Access* **2021**, *9*, 62867–62879. [\[CrossRef\]](#)
28. Kim, H.; Jin, W.; Lee, H. MmWave Path Loss Modeling for Urban Scenarios Based on 3D-Convolutional Neural Networks. In Proceedings of the 2022 International Conference on Information Networking (ICOIN), Jeju, Korea, 12–15 January 2022.
29. Wu, L.; He, D.; Ai, B.; Wang, J.; Liu, D.; Zhu, F. Enhanced Path Loss Model by Image-Based Environmental Characterization. *IEEE Antennas Wirel. Propag. Lett.* **2022**, *21*, 903–907. [\[CrossRef\]](#)
30. Ratnam, V.V.; Chen, H.; Pawar, S.; Zhang, B.; Zhang, C.J.; Kim, Y.J.; Lee, S.; Cho, M.; Yoon, S.R. FadeNet: Deep Learning-Based Mm-Wave Large-Scale Channel Fading Prediction and Its Applications. *IEEE Access* **2021**, *9*, 3278–3290. [\[CrossRef\]](#)
31. Zhang, X.; Shu, X.; Zhang, B.; Ren, J.; Zhou, L.; Chen, X. Cellular network radio propagation modeling with deep convolutional neural networks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 6–10 July 2020; pp. 2378–2386.
32. Hayashi, T.; Nagao, T.; Ito, S.A. A Study on the variety and size of input data for radio propagation prediction using a deep neural network. In Proceedings of the 2020 14th European Conference on Antennas and Propagation (EuCAP), Copenhagen, Denmark, 15–20 March 2020.
33. Inoue, K.; Ichige, K.; Nagao, T.; Hayashi, T. Radio Propagation Prediction Using Deep Neural Network and Building Occupancy Estimation. *IEICE Commun. Express* **2020**, *9*, 506–511. [\[CrossRef\]](#)
34. SVR PATHLOSS. Available online: https://github.com/timotrob/SVR_PATHLOSS (accessed on 2 January 2022).
35. LoRaWAN Measurement Campaigns in Lebanon. Available online: <https://zenodo.org/record/1560654#X-9B-VUzbIV> (accessed on 2 January 2022).
36. Path Loss Prediction. Available online: <https://github.com/lamvng/Path-loss-prediction> (accessed on 2 January 2022).
37. Timoteo, R.D.A.; Cunha, D.C.; Cavalcanti, G.D.C. A Proposal for Path Loss Prediction in Urban Environments Using Support Vector Regression. In Proceedings of the The Tenth Advanced International Conference on Telecommunications, Paris, France, 20–24 July 2014; pp. 119–124.
38. El Chall, R.; Lahoud, S.; El Helou, M. LoRaWAN Network Radio Propagation Models and Performance Evaluation in Various Environments in Lebanon. *IEEE Internet Things J.* **2019**, *6*, 2366–2378. [\[CrossRef\]](#)
39. Popoola, S.I.; Atayero, A.A.; Arausi, O.D.; Matthews, V.O. Path Loss Dataset for Modeling Radio Wave Propagation in Smart Campus Environment. *Data Br.* **2018**, *17*, 1062–1073. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Mapbox. Static Maps. 2022. Available online: <https://www.mapbox.com/static-maps> (accessed on 18 January 2022).
41. Li, L.; Jamieson, K.; DeSalvo, G.; Rostamizadeh, A.; Talwalkar, A. Hyperband: A Novel Bandit-Based Approach To Hyperparameter Optimization. *J. Mach. Learn. Res.* **2018**, *18*, 6765–6816.
42. Horn, Z.C.; Auret, L.; McCoy, J.T.; Aldrich, C.; Herbst, B.M. Performance of Convolutional Neural Networks for Feature Extraction in Froth Flotation Sensing. *IFAC-PapersOnLine* **2017**, *50*, 13–18. [\[CrossRef\]](#)
43. Wang, S.; Han, K.; Jin, J. Review of Image Low-Level Feature Extraction Methods for Content-Based Image Retrieval. *Sens. Rev.* **2019**, *39*, 783–809. [\[CrossRef\]](#)
44. Paramkusham, S.; Rao, K.M.M.; Rao, B.F. Comparison of Rotation Invariant Local Frequency, LBP and SFTA Methods for Breast Abnormality Classification. *Int. J. Signal Imaging Syst. Eng.* **2018**, *11*, 136–150. [\[CrossRef\]](#)

45. Choras, R.S. Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems. *Int. J. Biol. Biomed. Eng.* **2007**, *1*, 6–15.
46. Sukiman, T.S.A.; Suwilo, S.; Zarlis, M. Feature Extraction Method GLCM and LVQ in Digital Image-Based Face Recognition. *Sinkron* **2019**, *4*, 1–4. [\[CrossRef\]](#)
47. Casagrande, L.; Macarini, L.A.B.; Bitencourt, D.; Fröhlich, A.A.; de Araujo, G.M.A. New Feature Extraction Process Based on SFTA and DWT to Enhance Classification of Ceramic Tiles Quality. *Mach. Vis. Appl.* **2020**, *31*, 71. [\[CrossRef\]](#)
48. Althnani, A.; Aloboud, N.; Alkharashi, N.; Alduwaish, F.; Alrshoud, M.; Kurdi, H. Face Gender Recognition in the Wild: An Extensive Performance Comparison of Deep-Learned, Hand-Crafted, and Fused Features with Deep and Traditional Models. *Appl. Sci.* **2021**, *11*, 89. [\[CrossRef\]](#)
49. Ghazali, K.H.; Mansor, M.F.; Mustafa, M.M.; Hussain, A. Feature extraction technique using discrete wavelet transform for image classification. In Proceedings of the 2007 5th Student Conference on Research and Development, Selangor, Malaysia, 11–12 December 2007.
50. Costa, A.F.; Humpire-Mamani, G.; Traina, A.J.M.H. An Efficient Algorithm for Fractal Analysis of Textures. In Proceedings of the Brazilian Symposium on Computer Graphics and Image Processing, Ouro Preto, Brazil, 22–25 August 2012; pp. 39–46.
51. Yamashita, R.; Nishio, M.; Do, R.K.G.; Togashi, K. Convolutional Neural Networks: An Overview and Application in Radiology. *Insights Imaging* **2018**, *9*, 611–629. [\[CrossRef\]](#)
52. Ranjbar, S.; Nejad, F.M.; Zakeri, H.; Gandomi, A.H. Computational intelligence for modeling of asphalt pavement surface distress. In *New Materials in Civil Engineering*; Elsevier: Oxford, UK, 2020.
53. Weimer, D.; Scholz-reiter, B.; Shpitalni, M. Design of Deep Convolutional Neural Network Architectures for Automated Feature Extraction in Industrial Inspection. *CIRP Ann. Manuf. Technol.* **2016**, *65*, 417–420. [\[CrossRef\]](#)
54. Wang, L.; Wang, X.; Chen, A.; Jin, X.; Che, H. Prediction of Type 2 Diabetes Risk and Its Effect Evaluation Based on the XGBoost Model. *Healthcare* **2020**, *8*, 247. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Wang, H.; Liu, C.; Deng, L. Enhanced Prediction of Hot Spots at Protein-Protein Interfaces Using Extreme Gradient Boosting. *Sci. Rep.* **2018**, *8*, 14285. [\[CrossRef\]](#)
56. Niu, Y. Walmart Sales Forecasting Using XGBOOST Algorithm and Feature Engineering. In Proceedings of the 2020 International Conference on Big Data and Artificial Intelligence and Software Engineering (ICBASE), Bangkok, Thailand, 30 October–1 November 2020.
57. Adebayo, S. How The Kaggle Winners Algorithm XGBoost Algorithm Works. Available online: <https://dataaspirant.com/xgboost-algorithm/> (accessed on 16 February 2022).
58. Xia, Y.; Liu, C.; Li, Y.Y.; Liu, N.A. Boosted Decision Tree Approach Using Bayesian Hyper-Parameter Optimization for Credit Scoring. *Expert Syst. Appl.* **2017**, *78*, 225–241. [\[CrossRef\]](#)
59. Xgboost Developers. *Release 1.5.0-Dev Xgboost Developers*; Technical Report; Xgboost Developers: Seattle, WA, USA, 2021.
60. Nagao, T.; Hayashi, T. Study on Radio Propagation Prediction by Machine Learning Using Urban Structure Maps. In Proceedings of the 14th European Conference on Antennas and Propagation (EuCAP2020), Copenhagen, Denmark, 15–20 March 2020.
61. Chen, T.; Guestrin, C.X.A. Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016.
62. Rashmi, K.V.; Gilad-Bachrach, R.D. Dropouts Meet Multiple Additive Regression Trees. *J. Mach. Learn. Res.* **2015**, *38*, 489–497.
63. Ojo, S.; Imoize, A.; Alienyi, D. Radial Basis Function Neural Network Path Loss Prediction Model for LTE Networks in Multitransmitter Signal Propagation Environments. *Int. J. Commun. Syst.* **2021**, *34*, e4680. [\[CrossRef\]](#)
64. Ebhota, V.C.; Isabona, J.; Srivastava, V.M. Investigating Signal Power Loss Prediction in A Metropolitan Island Using ADALINE and Multi-Layer Perceptron Back Propagation Networks. *Int. J. Appl. Eng. Res.* **2018**, *13*, 13409–13420.
65. Moraitis, N.; Vouyioukas, D.; Gkioni, A.; Louvros, S. Measurements and Path Loss Models for a TD-LTE Network at 3.7 GHz in Rural Areas. *Wirel. Netw.* **2020**, *26*, 2891–2904. [\[CrossRef\]](#)
66. Aldosary, A.M.; Aldossari, S.A.; Chen, K.C.; Mohamed, E.M.; Al-Saman, A. Predictive Wireless Channel Modeling of Mmwave Bands Using Machine Learning. *Electronics* **2021**, *10*, 3114. [\[CrossRef\]](#)
67. Moraitis, N.; Tsipi, L.; Vouyioukas, D. Machine learning-based methods for path loss prediction in urban environment for LTE networks. In Proceedings of the International Conference on Wireless and Mobile Computing, Networking and Communications, Thessaloniki, Greece, 12–14 October 2020.
68. Ojo, S.; Sari, A.; Ojo, T.P. Path Loss Modeling: A Machine Learning Based Approach Using Support Vector Regression and Radial Basis Function Models. *Open J. Appl. Sci.* **2022**, *12*, 990–1010. [\[CrossRef\]](#)
69. Garcia, M.V. Interpretable Forecast of NO2 Concentration Through Deep SHAP. Master's Thesis, Universidad Nacional de Educacion a Distancia, Madrid, Spain, 2019.
70. Jimoh, A.A.; Bakinde, N.T.; Faruk, N.; Bello, O.W.; Ayeni, A.A. Clutter Height Variation Effects on Frequency Dependent Path Loss Models at UHF Bands in Build-Up Areas. *Sci. Technol. Arts Res. J.* **2015**, *4*, 138–147. [\[CrossRef\]](#)
71. Kuhn, M.; Johnson, K. *Feature Engineering and Selection: A Practical Approach for Predictive Models*; CRC Press: New York, NY, USA, 2020.
72. Molnar, C. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*; Lulu Press Inc.: Morrisville, NC, USA, 2022.