



Article

Predictive Wireless Channel Modeling of MmWave Bands Using Machine Learning

Abdallah Mobark Aldosary¹, Saud Alhajaj Aldossari^{2,*} , Kwang-Cheng Chen³, Ehab Mahmoud Mohamed^{2,4} and Ahmed Al-Saman^{5,*} 

¹ Department of Computer Science, Prince Sattam bin Abdulaziz University, As Sulayyil 18211, Saudi Arabia; ab.aldosary@psau.edu.sa

² Department of Electrical Engineering, Prince Sattam bin Abdulaziz University, Wadi Addwasir 11991, Saudi Arabia; e.mahmoud@psau.edu.sa

³ Department of Electrical Engineering, University of South Florida, Tampa, FL 33620, USA; kwangcheng@usf.edu

⁴ Electrical Engineering Department, Aswan University, Aswan 81542, Egypt

⁵ Faculty of Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway

* Correspondence: s.alhajaj@psau.edu.sa (S.A.A.); ahmedsecure99@gmail.com (A.A.-S.)

Abstract: The exploitation of higher millimeter wave (MmWave) is promising for wireless communication systems. The goals of machine learning (ML) and its subcategories of deep learning beyond 5G (B5G) is to learn from the data and make a prediction or a decision other than relying on the classical procedures to enhance the wireless design. The new wireless generation should be proactive and predictive to avoid the previous drawbacks in the existing wireless generations to meet the 5G target services pillars. One of the aspects of Ultra-Reliable Low Latency Communications (URLLC) is moving the data processing tasks to the cellular base stations. With the rapid usage of wireless communications devices, base stations are required to execute and make decisions to ensure communication reliability. In this paper, an efficient new methodology using ML is applied to assist base stations in predicting the frequency bands and the path loss based on a data-driven approach. The ML algorithms that are used and compared are Multilayers Perceptrons (MLP) as a neural networks branch and Random Forests. Systems that consume different bands such as base stations in telecommunications with uplink and downlink transmissions and other internet of things (IoT) devices need an urgent response between devices to alter bands to maintain the requirements of the new radios (NR). Thus, ML techniques are needed to learn and assist a base station to fluctuate between different bands based on a data-driven system. Then, to testify the proposed idea, we compare the analysis with other deep learning methods. Furthermore, to validate the proposed models, we applied these techniques to different case studies to ensure the success of the proposed works. To enhance the accuracy of supervised data learning, we modified the random forests by combining an unsupervised algorithm to the learning process. Eventually, the superiority of ML towards wireless communication demonstrated great accuracy at 90.24%.

Keywords: beyond 5G; mmWave; wireless channel modeling; bands; path loss; machine learning; MLP; random forests



Citation: 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. <https://doi.org/10.3390/electronics10243114>

Academic Editor: Sotirios K. Goudos

Received: 7 October 2021

Accepted: 2 December 2021

Published: 14 December 2021

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



Copyright: © 2021 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

Machine Learning deployments in wireless communication systems is rapidly gaining attractiveness with the usage of data-driven approaches [1]. The justification of that is the usage of the availability of the data solutions to complex and high computational machines for wireless communication problems [2]. Furthermore, the new emergence of the Quantum Computing (QC) technologies has a strong potential to be applied to the future wireless generations with the deployment of ML methods [3]. The potential investigations of ML usage in the field of wireless communications has been increased, especially in channel

modeling. Initiation studies of applying ML to the general discipline of wireless networks can be found in [4]. The expectations of deploying ML techniques in the B5G will be a major part of designing the networks to be more autonomous, dynamic and self-organizing [5] used deep learning methods such as the multiple layers perceptrons (MLP) algorithm to predict the wireless channel parameters for dedicated short-range communication and were able to outperform formal empirical models. Furthermore, other ML algorithms have been used in [6] to assist BS to predict the acceptable received signal in millimeter wave (mmWave) bands, i.e., 30–300 GHz, based on classification methodologies [7] used ML to classify and identify the wireless channel Line of Sight (LoS) and non Line of Sight (NLOS). Applying ML techniques towards wireless channel modeling seems successful usage [8]. However, these methods performance relies on data [9] and to avoid the complexity of the wireless channel characterization, a crucial ML methodologies have to be involved.

One fact of the wireless channel modeling is that the higher frequency bands usage such as mmWave bands, the higher signal's energy degradation. This fact leads to change the behavior of the signal with higher mmWave frequencies, where the propagated signal becomes more sensitive during the communication leading to an increase in the path loss. Moreover, penetration of the propagated signal through objects is less in the higher mmWave frequencies [10]. Thus, investigating the wireless channel modeling in higher frequencies is crucial for B5G. The authors in [11] investigated the D-band for higher frequency bands and presented indoor scenarios and recommended investigating longer distances to enhance the coverage area. Other authors have used ML techniques to predict the path loss exponent and shadowing factor using satellite images such as [12].

The main goal of the current study is to assist the base stations to predict the channel state information such as frequency bands and the path loss using ML techniques, for instance artificial neural networks (ANN) and Random Forest. The wireless communications scenario will be an urban macro environment and the frequency band will be 1–100 GHz to investigate the bands, and another dataset will be utilized to validate our methodology for varieties of mmWave frequencies. The construction and development of the wireless channel modeling with the usage of the ML is becoming important in the new radio (NR). Since path loss is the reduction of power strength during signal propagation through a channel, it's becoming a key factor with the Mmwave journey. The Federal Communications Commission (FCC) has already assigned the MmWave band in the United States as 24, 28, 37, 39, 47 GHz for the 5G and selected bands above frequency 95 GHz for experimental usage. Furthermore, the FCC is encouraging industry and academia to start investigating the higher MmWave frequencies [13]. Researchers have already started performing case studies towards higher bands such as 28 GHz in [14]. Thus, we formulate this work to meet this request by investigating the higher MmWave bands. Furthermore, the major goal of this manuscript is to investigate the path loss and bands in high MmWave frequencies for B5G other than the traditional procedure to pave the road towards future generations as shown in Figure 1. Another goal is to assist base stations in predicting the path loss and the frequency band that will result in enhancing the communications.

Concerning classical works, [15] presented a new measurement campaign to measure mmWave channel using classical empirical measurements. The classical procedure can be obtained using empirical or deterministic models. The empirical models are based on applying statistical methods to data that have been collected via the measurement filed. The disadvantage with this approach is that it is time consuming because every environment has to be measured to obtain an inaccurate path loss model. While the deterministic models are based on geometrical theories of electromagnetic propagation, such as ray-tracing, which is accurate but requires more details of the surrounding environment, and is therefore complex [16]. The authors of [17] have proposed a method to reduce the complexity of the channel estimation using an extended Kalman filter (EKF) and unscented Kalman filter (UKF) in a vehicle to vehicle (V2V) and IoT. Ref. [18] proposed a new methodology to predict the path loss of a wireless channels were based on a dataset from a different environment where ML algorithms were applied to learn the wireless medium

pattern and predict the loss with perfect accuracy. This paper proposes ML techniques to assist the base station to predict frequency bands and path loss as recommended by [19] to involve ML due to the complex heterogeneous nature of the network structures and wireless services. The dataset was used to train the neural network's method to predict the channel state information (CSI) features in the higher bands and how these features are preformed compared to the lower bands. Deep learning neural networks methods and other ML techniques such as Random Forest are used to assist the base station to predict the frequency bands, path loss, and how a higher mmWave band affects the wireless channel modeling. The deep learning technique will be Multilayer Perceptrons (MLP), since it works better than other artificial neural network algorithms with non-stationary data. Forward and backward problem propagation and error correction methods will be used during the testing phase. Moreover, optimization algorithms that deal with optimizing the neural networks will be explained in the following sections. To enhance the accuracy of the prediction models, the unsupervised learning technique is involved in supervised learning to obtain the CSI features patterns and denominational reduction.

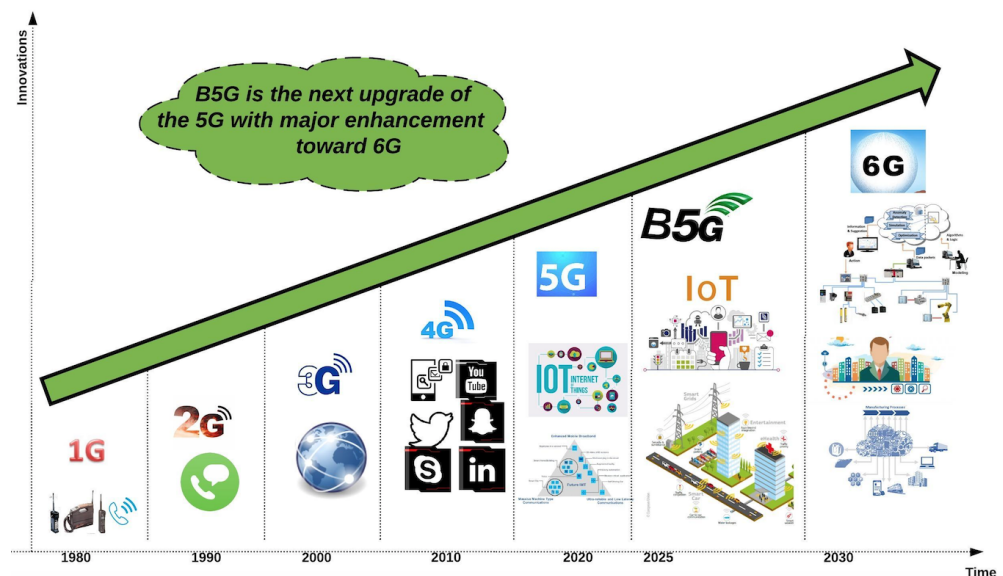


Figure 1. Evolution of cellular network generations.

This work is promising for future wireless systems since each standard has its own regulations. One of these regulations in wide area networks (WAN) such as cellular networks or Wireless Local Area Network (WLAN), is assigning the frequency bands. For instance, the proposed methodology can assist a base station in predicting the LTE and 5G frequency bands or the 802.11g systems that use 2.4 GHz and the 802.11a system that uses 5 GHz. The newest Wi-Fi system is called 802.11n and it uses both 2.4 GHz and 5 GHz. The proposed scheme can be further implemented towards Wireless Sensor Networks (WSN) or even further to the Personal Area Network (PAN) systems. Moreover, other applications of this scheme are telecommunication companies where each cellular company has its own bands. Since the new devices are SIMless, new devices are capable of predicting the frequency bands that belong to that specific company. Thus, work is promising for future regulations and standards in future wireless generations. Therefore, by implementing the proposed work, future wireless devices should not be specified for a specific frequency band.

The rest of this paper is arranged as follows. Section 1 is an introduction of the wireless generations and ML. Section 2 gives system model formulation, while Section 3 presents data driven and preprocessing. Section 4 elaborates on the proposed supervised learning methods for B5G. Results and discussion are presented in Section 5 while the following section is the conclusion.

2. System Model Formulation

This paper proposes a new scheme to assist the base station to predict a wide range of frequency bands and path loss. In a practical scenario, however, the receiver has no access to the actual bands, and is supposed to predict it. The proposed system is a single-input single-output (SISO) wireless communication system with a macro urban environment. The system has a radio frequency bandwidth of 800 MHz with a transmitter power of 30 dBm, and the transmit and receive array type are URA. Due to the weakness of higher MmWave bands, other effects related to weather have been considered, such as barometric pressure 1013.25 mbar, humidity 50%, the temperature is 20 degrees Celsius, and rain rate was assumed to be zero [20]. Other system parameters can be summarized in Table 1.

Table 1. Channel measurement parameters.

Parameters	Values
Distance (m)	1–300
Frequency (GHz)	95
Bandwidth (MHz)	800
TXPower (dBm)	300
Scenario	UMa
Polarization	Co-Pol
TxArrayType	ULA
RxArrayType	ULA
Antenna	SISO
Tx/Rx antenna Azimuth and Elevation (red)	10

Where it exhibits that the channel measurement parameters of the data raw that was used for this paper. Python was used to perform the data analysis. A flowchart of the model or the scenario of this manuscript can be seen in Figure 2.

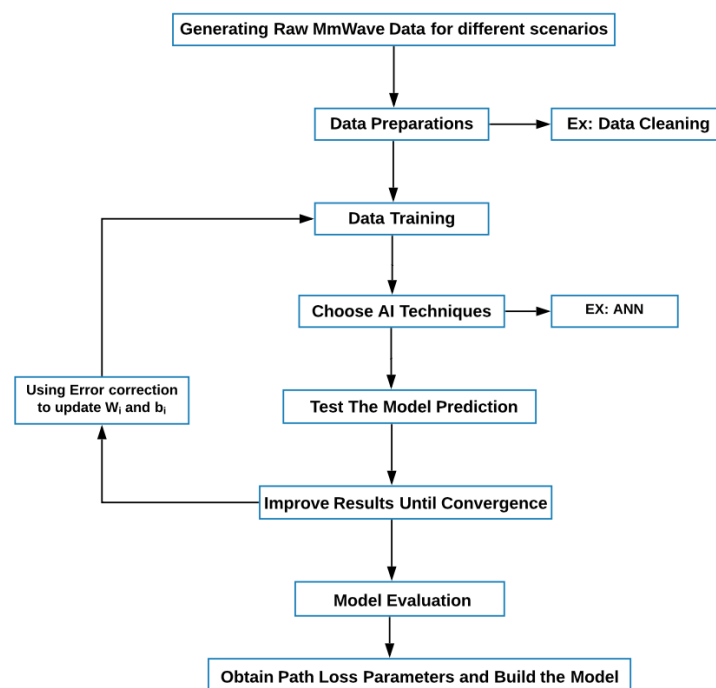


Figure 2. Flowchart of the proposed manuscript.

Theorem

Investigating the effects of higher mmWave can be driven with the usage of the Friis Theorem [21] by assuming an environment with only a transmitter T_x and a receiver T_r and no obstacles between them to create free space. The distance is d and a transmitted power of P_t and omnidirectional antennas at both T_x and T_r . Therefore, the received power P_r is calculated using

$$P_r(dBm) = \frac{P_t}{4\pi d^2} G_t \quad (1)$$

here G_t is the gain of the transmit antenna. Assuming the antenna at the receiver side has an effective aperture A_{ER} . Then (1) becomes.

$$P_r(dBm) = \frac{P_t}{4\pi d^2} G_t A_{ER} \quad (2)$$

The effective aperture of the antenna can be:

$$A_{ER} = \frac{\lambda^2}{4\pi} G_r \quad (3)$$

$$P_r(dBm) = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} \quad (4)$$

where $\lambda = \frac{c}{f_c}$, λ is the wavelength in meters, c is the speed of light and f_c is the frequency in GHz [22]. Moreover, Friis theory in (4) can lead to the same conclusion that path loss is proportional to the frequency bands.

$$P_r(dBm) = \frac{P_t G_t G_r c^2}{(4\pi f d)^2} \quad (5)$$

From (5), we can infer the frequency bands as shown below.

$$f(MHz) = \frac{c^2}{4\pi d} \sqrt{\frac{P_t G_t G_r}{P_r}} \quad (6)$$

The above equation proves that with increasing the frequency, the received power reduced which means the path loss will increase as well as shown in the below equations.

$$PL(dBm) = 10 \log_{10} \left(\frac{16\pi^2 d^n}{\lambda^2} \right) \quad (7)$$

$$P_r(d) = P_t \frac{G_t G_r \lambda}{(2 * \pi)^2 d^n L} \quad (8)$$

Understanding the behavior of the path loss (L) concerning distance and other CSI features with continuous dependent variables leads to the use of the proper procedure of the ML categories. Therefore, it is clear in our case that the ML strategy will be a supervised regression problem based on a data-driven approach. One technique to measure the energy loss is by using the path loss model, which measures the reduction of the energy during the propagation. By using path loss models, energy loss can be governed. The complexity and the accuracy of the path loss models can vary with many factors, such as the environment, interference level, energy, distance, and so on. The wireless bands have a high impact on the amount of loss and coverage following the equation of $f = \frac{c}{\lambda}$ that is different from the lower bands. CSI features that are used in our system are frequency bands (GHz), T-R separation distance (m), received power (dBm), phase (rad), azimuth AoD (degree), elevation AoD (degree), azimuth AoA (degree), elevation AoA (degree), path Loss (dB) and root mean square (RMS) delay spread (ns). Frequency band (GHz) and path Loss (dB) will be used as targets where other CSI will be used to feed the NNs to assist the base station to predict the bands or the path loss.

In the macro environment, user equipment (UEs) are assumed to be non-stationary and uniformly distributed. In our model, we applied deep learning NNs techniques such as MLP as supervised learning to predict the dependent variable. Having a dataset consisting of the dependent variable y_i and CSI x_i , features of the mmWave bands can be predicted. In this paper, we will focus on predicting the path loss and the frequency bands based on previous data $p(y/x)$. The data are split into training $D = (x_i, y_i)_{i=1}^N$ and testing $T = (x_i, y_i)_{i=1}^N$ where $D \cap T = \phi$.

The ML algorithms learn the pattern and assist base stations to make a decision due to the complexity of the network structures and wireless services that are recommended for the new radio techniques [23]. ANNs are widely used nowadays to learn the complex pattern of the wireless channel to avoid complex and unreliable mathematical formulations. Since the presented data is nonlinear and multivariable characteristics, ANN could be involved to predict the frequency bands and path loss to assist the base stations as alternative model structures for received power. The ANN's structure consists of at least three main layers, input (i), hidden (j_n), and output (k) layers, and for simplification, we assume the system only consists of only one hidden layer. Each layer composes of one or more number of neuron n_{li} , where l to represent which layer and $i \in \mathbf{n} = (1, 2, \dots, k)$ for identify the specific neurons which considered the main component and the processing unit. n_k neurons in layer j_n feeds from n_{k-1} neurons in the previous layer by a weight vector. Input of CSI $X = (x_1, x_2, \dots, x_n)$ are fed to networks and then multiplied by the weight vector $W = (w_1, w_2, \dots, w_n)$ and with addition of bias variable b_i then summed at the hidden layer. Then, activation function $f(\cdot)$ is used for every node in the hidden layers to produce an output where more details of the activation functions will be elaborated later in this manuscript. The summation of the hidden and output layer of the NNs denoted by s_j and s_k respectively and can be formulated as follows.

$$s_j = \sum_{i=0}^k w_{ij} \cdot x_i + b_i \quad (9)$$

$$s_k = \sum_{i=0}^k w_{jk} \cdot x_k + b_i \quad (10)$$

$$y = f\left(\sum_{i=1}^n w_i x_i + b_i\right) \quad (11)$$

Figure 3 is a simplified systems model, where the training data samples $D = (x_i, y_i)_{i=1}^N$ are feed into the system. Forward, backward, and optimization techniques are implemented in this system. Deep learning algorithms will be compared with ML methods and then optimization techniques will be applied to reduce the loss of the prediction.

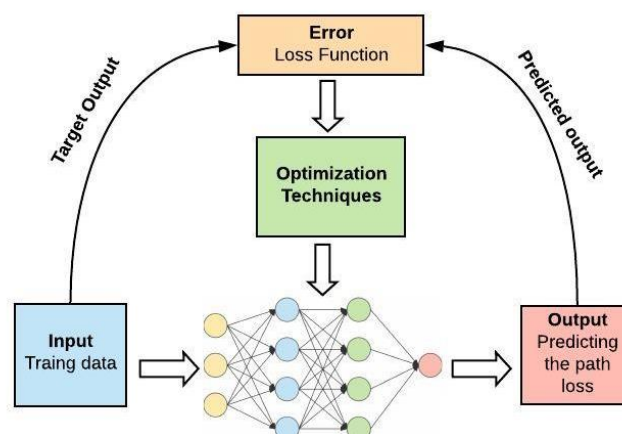


Figure 3. A general visualization explaining of the system model using the ANN technique.

3. Data Driven Modeling and Preprocessing

Data-driven is considered as a new approach for next-generation generation models [24]. Data in the proposed cases is produced using the NYUSIM simulator, which is based on an extensive measurement campaigns for varieties of mmWave bands and different scenarios [25]. After the data has been produced, data processing has to be performed, such as data cleansing. Data cleansing is one of the main parts of machine learning that perform a significant part to enhance the accuracy of the model. Data cleansing is simply the process of removing erroneous or unwanted observations from the data and then replacing them with samples based on several concepts such as the averaging. Furthermore, it involves managing unwanted outliers and then verifying the data and removing unwanted observations including detecting redundant or irrelevant samples from the dataset [26]. The dataset of CSI features has been divided into training, testing, and validating with percentages of 70%, 15%, and 15%, respectively. Moreover, in this section, two methods of ML will be investigated to obtain the optimum results. The techniques are Random Forest and MLP as approaches to NNs.

3.1. Channel Impulse Response

The wireless channel can be modelled differentially depending on the communication environments. Linear time-variant (LTV) channel model is a popular wireless channel model $h(t)$ and in theory, if the received signal $r(t)$ with an input signal $s(t)$

$$r(t) = x(t) * h(t) + n(t) \quad (12)$$

where the $n(t)$ is noise that affects propagated signal and the convolution is given by integration with a channel delay τ as shown below.

$$r(t) = \int_{-\infty}^{\infty} x(\tau)h(t - \tau)d\tau + n(t) \quad (13)$$

$$h(t) = \sum_{i=0}^{N-1} a_i e^{j\theta_i} \delta(t - \tau_i) \quad (14)$$

$$APDP(\tau) = \frac{1}{N} \sum_{i=0}^N |h_i(\tau)|^2 \quad (15)$$

Equation (14) is the channel impulse response of the wireless channel where N is the multipath components (MPCs) and a_i is the random amplitude of MPC which can be characterized statistically using Rayleigh or Rician distribution based on the communication environment. Then, the power delay profile can be obtained by taking the squared absolute value of the impulse response. APDP (tau) in (15) represents the average power delay profile (PDP) that involves removing the effects of the small scale fading by averaging the power delay profile. PDP usually gives the intensity of the received signal of a multipath channel as a function of time delay [27].

$$\begin{aligned} PDP &= |h(t)|^2 \\ &= \sum_{i=0}^{N-1} (a_k e^{j\theta_i})^2 \delta(t - \tau_k) \end{aligned} \quad (16)$$

However, the predicted path loss can be obtained by the summation of the predicted channel impulse response as following [28].

$$\hat{P}L = \frac{1}{N} \sum_{i=0}^{N-1} \hat{h}_i(\mathbf{x}) \quad (17)$$

$$PL[dBm] = 10 \log_{10} \left(\frac{P_t}{P_r} \right) = 10 \log_{10} \left(\frac{4\pi d}{\lambda} \right)^2 \quad (18)$$

In the following subsections, multiple ML and DL techniques will be used to investigate the wireless channel of the mmWave band for beyond 5G. Furthermore, applying these techniques to the base station to predict the frequency bands and the path loss. The ML techniques establish the mapping relationships between the dependent and independent variables. In our case, the dependent variable is the path loss and bands while the other CSI represents the independent variables.

3.2. Random Forest

The main idea of Random Forests is aggregating nearly random generalizations to build a strong decision as supervised learning. Random Forests is a ML method that is based on ensemble learning and can be applied to regression and classification tasks [29]. Random Forests use Bootstrap Aggregation to divide the data into multiple bags and each bag passes and trains the data to the decision tree as shown in Algorithm 1. This method is called bagging and the Random Forest is an extension of the bagging technique. The bagging is implemented by selecting the subset of the data arbitrarily and selecting the features randomly based instead of using all features. The estimator trees are combined to determine the output instead of relying on an individual decision. Thus, for predicting the frequency bands to assist the base station, the final prediction value is obtained by averaging the whole predictions from all single trees as shown in (19) [9]. Ensemble techniques are used to enhance the prediction of an ML model, and there are two approaches of ensemble learning, as outlined below.

- Bagging: Many estimators are built independently using a subset of the training data and with the usage of the data set, then prediction outputs are averaged. The bagging method reduces the variance and low impact on the bias. The learning processes are in parallel and an example of it is the Random Forests.
- Boosting: This is a supporting mechanism where estimators are built sequentially. This means that the learning process builds of an estimator is based on prior learning of a different estimator. Boosting is based on combining multiples of weak learners to form a stronger one and an example of it is adaptive boosting (known as Adaboost).

Bagging Regressor is an ensemble combiner estimator that is used to fit a subset of the original dataset and aggregate (combine) the prediction of each estimator by either averaging final prediction for regression or voting for classification techniques. This method is used to reduce the variance and the bias of a black box estimator since every single decision tree exhibits high variance and tends to be overfitting. Another feature of the Random Forests technique is reducing the correlation between the sampled trees. The main difference between the bagging and the pasting techniques is the replacement of the randomly drawn subset that contains multiple features where bagging does the replacement while the pasting does not reuse the drawn samples. The disadvantages of Random Forests are the computational complexity compared to the decision tree and their non-interpretability.

The Random Forests are implemented based on a regression approach to predict the dependent variables such as frequency bands and path loss. The input to that learning is the CSI features in the system formulation, and the number of estimators tree is 100 based on the rule of thumb. The number of features that a node of the tree estimator is set to be max features = 7.

Since the dataset has 11 features, the procedure of implementing the Random Forest technique can be seen as follows [30].

To enhance the Random Forest regression prediction, Principal Component Analysis (PCA) technique, which is unsupervised learning, is implemented to find more patterns to enhance the prediction.

$$\hat{y}_b = \frac{1}{T} \sum_{t=1}^T \hat{d}_t(x) \quad (19)$$

Algorithm 1: Random Forests Regressor

Input: A bootstrap subset of CSI dataset is drawn from the original training set with replacement $A \subset \mathbf{X}$

- 1 where $(x, y) \in A$.
- Output:** Compute model $\hat{y}_{rf}(x) = \frac{1}{T} \sum_{i=0}^T \hat{y}_i(x)$
- 2 Compute every decision tree in the forests using the above procedure.
- 3 **repeat**
- 4 **for** $(x, y)_i \in \mathbf{X}$ **do**
- 5 T_{i+1} regression tree is fitted to each subset;
- 6 **for** T_{node_i} nodes of the tree estimator **do**
- 7 select a portion of the features that specific subset
- 8 **end**
- 9 **end**
- 10 **until** stopping criterion is not met;

Another way to investigate the signal attenuation is through checking the path loss for different mmWave bands from 1 GHz through 100 GHz. As can be seen in the figure below, in specific bands, the path loss is considered and performed better than other bands.

Based on Figure 4, path loss (dBm) VS Bands (GHz) visualization explains the low loss through different frequency bands. We can identify the perfect path loss in bands such as 2.4, 17, 41, 54, 68, and 82 GHz. The number of estimators in this kind of ensemble method is the number of trees to be used in the forest. In our case, we used 100 estimators where each tree or estimator makes a prediction and with the other tree's prediction can be averaged to obtain the final prediction.

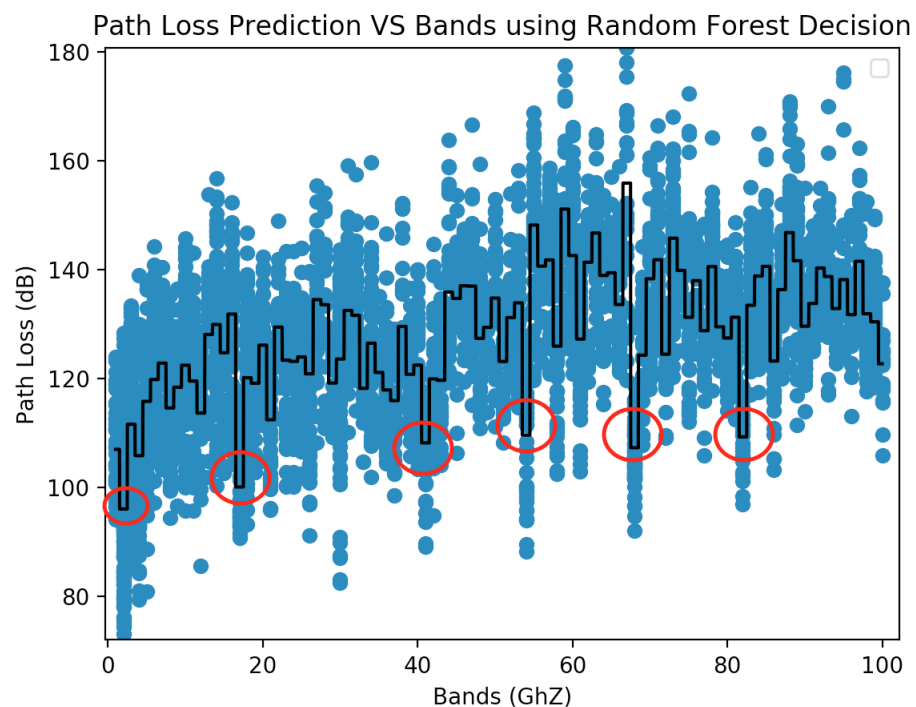


Figure 4. A general visualization of the high loss via different MmWave frequencies.

3.3. Multilayer Perceptrons

The Multilayer Perceptrons (MLP) method is a branch of artificial neural networks (ANN) for the purpose of computing and prediction. MLP combines neurons in different layers to solve complex problems similar to Figure 5. MLP will be used as a regression techniques to investigate the high MmWave where Feed-forward neural networks and a

back-propagation computation will be used to assist the prediction in the wireless base station. This section describes the proposed NN technique to investigate wireless channel modeling in terms of path loss and bands. The output of the neural networks will be fed back to the network to compute the backpropagation to reduce the loss using the following equation.

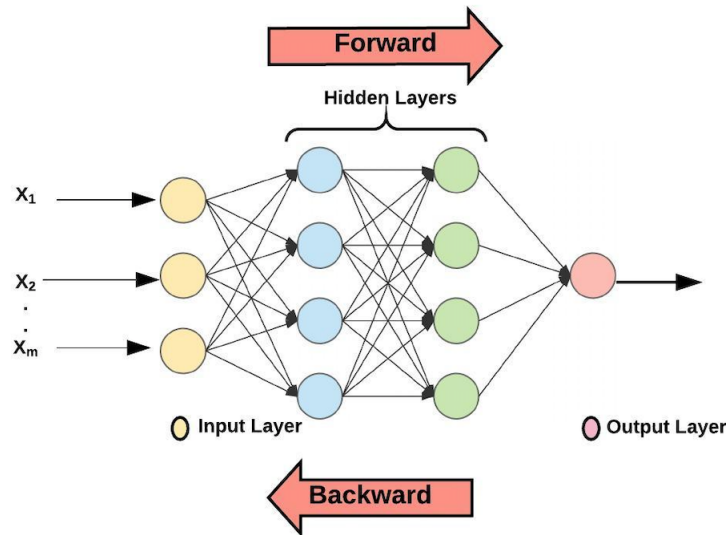


Figure 5. A general visualization explaining the MLP forward and backward algorithms.

$$\mathcal{L}_e(y, f(x, w)) = \frac{1}{2} \sum_{i=0}^k (y_{k_i} - \hat{y}_{k_i})^2 \tag{20}$$

where \mathcal{L} is the loss, y is the actual dependent variable and \hat{y}_k is predicted value. The predicted value of the neural networks can be seen as follows.

$$\hat{y}_k = \sum_{i=0}^k (w_{jk} \cdot S \sum_{i=0}^k (w_{ij}x_i + b_j) + b_k) \tag{21}$$

Then, optimization methods are used to reduce the error to reach the minima of loss. For simplification, the bias variable will be initialized to zero from now on. Since the loss has been obtained in (20), backpropagation feeds the parameter values back to the network to reduce the error then after some iterations, the parameter values are ready to be used to form the path loss model of 5G advance. Backpropagation calculations shall be started from the output layer toward the first hidden layer and by taking the derivative with some derivative properties such as the chain rule as given below.

$$\begin{aligned} \frac{\partial \mathcal{L}(w)}{\partial (w_{jk})} &= \frac{\partial \frac{1}{2} \sum_{i=0}^k (y - \hat{y})^2}{\partial w_{jk}} \\ &= (y - \hat{y}) \frac{\partial (y - \hat{y})}{\partial w_{jk}} \\ &= (y - \hat{y}) \frac{\partial \hat{y}}{\partial w_{jk}} \\ &= (y - \hat{y}) \frac{\partial \hat{y}}{\partial x_k} \cdot \frac{\partial x_k}{\partial w_{jk}} \\ &= (y - \hat{y}) \cdot A_{x_k} \cdot (1 - A_{x_k}) \cdot x_j \\ &= Error \cdot A_{x_k} \cdot (1 - A_{x_k}) \cdot x_j \end{aligned} \tag{22}$$

where A_{x_k} is the derivation of the activation function and x_i, x_j, x_k are unit value in input, hidden and output layer respectively.

The value of the derivation in the above equations is called the backpropagated error from the output layer which passes the output value to the hidden layer. This procedure has to come through every layer in the networks until it reaches the input layer, as shown in the below derivation, to obtain the final backpropagation value. Then the procedure of updating the weight and bias parameters begins and proceeds anew with a forward propagation to the networks.

$$\begin{aligned}
 \frac{\partial \mathcal{L}(w)}{\partial (w_{ij})} &= \frac{\partial \frac{1}{2} \sum_{i=0}^k (y - \hat{y})^2}{\partial w_{ij}} \\
 &= \sum_{i=0}^k (y - \hat{y}) \frac{\partial \hat{y}}{\partial w_{ij}} \\
 &= \sum_{i=0}^k (y - \hat{y}) \frac{\partial \hat{y}}{\partial x_k} \cdot \frac{\partial x_k}{\partial x_j} \cdot \frac{\partial x_j}{\partial x_i} \cdot \frac{\partial x_i}{\partial w_{ij}} \\
 &= \sum_{i=0}^k (y - \hat{y}) \alpha(x_k) (1 - \alpha(x_k)) w_{jk} \alpha(x_j) (1 - \alpha(x_j)) x_i
 \end{aligned} \tag{23}$$

During the MLP regression procedure, optimization algorithms have been used to optimize the square loss. Gradients are usually used to find the minimum loss using a vector that reaches to the optimum loss. We started with the Broyden–Fletcher–Goldfarb–Shanno algorithm (LBFGS) optimizer which is a family member of quasi newton methods. We then started comparing the results with stochastic gradient descent (SGD). Moreover, another stochastic gradient-based optimizer has to be used that is called “adam” or sometimes referred to as adaptive moment estimation. Adam is an extension of stochastic gradient descent to update NNs’ weights iterative based in the training dataset [31]. The reason most of this work focused on the adam optimizer is that our data is non-stationary and it does not conserve a specific learning rate for every single weight update [32].

3.4. Neural Network Optimization

Optimization algorithms are often used in NNs such as Gradient Descent (GD), which is an iterative technique to obtain the values of the parameters of a function to reduce the cost function. Optimization algorithms are methods used to reduce the error function $E(x)$ that depends on the internal learning variables that are considered undependable to compute the response variable y . Thus, the main purpose of the optimization is to estimate the error gradient of the current state, then update the weights and the bias of the model using the back propagation of errors algorithm as shown in Algorithm 2. that explain the MLP with SGD that use back propagation to reduce the error. Other stochastic methods can be used to update the NNs, but there must be a trade-off between the precision of the updated parameters and the running time to update. Optimization methods are used to minimize the loss function $L(w_i, b_i)$ by finding the optimum wight w_i and b_i bias variables.

Optimization can be categorized into two major codes based on first or second-order optimization to obtain the information of the slope as a vector to identify the minima direction. The first-order optimization algorithms are used to minimize or maximize the loss function using a derivative. An example of the first-order methods is Gradient Descent (GD), where the derivative is used to identify whether the loss function is increasing or decreasing toward the minima. Second-order optimization algorithms are considered as optimizing the optimizer based on the second-order derivative to check how the gradient varies which usually involves a high computational cost. To avoid this dilemma, techniques such as LBFGS are applied with limited memory usage. Other SGD methods such as Adaptive Moment Estimation (Adam), Momentum, and Root Mean Square Propagation (RMSProp) are used as second-order optimization methods. Gradient Descent is considered

as the main optimizer to optimize an intelligent system. GD is used to reduce the loss by updating the model's parameters to reach the converge limit as shown in Algorithm 2. Once the gradient vector $\nabla J(W)$ is calculated, the new parameter can be obtained by multiplying a learning rate η by the gradient parameter and then subtracting them from the current weight parameter. Where the subtracting value is used to move toward the opposite direction since the gradient is considered as a vector that points to the direction of the increase of the loss function. The newly updated parameters are feedback to the network to reduce the cost. Where the cost can be calculated using the mean squared error. The new update parameter W_{t_1}

$$W_{t_1} = W_t - \eta \cdot \nabla J(W) \quad (24)$$

$$W_{t+1} = W_t - \eta \cdot J(W) \quad (25)$$

$$W_{t+1} = W_t - \eta \cdot \frac{\partial J(W)}{\partial w_t}$$

where W_{t+1} is a new weight that we assume will feed to the networks to reduce the loss and W_t is the previous weight parameter. Initialization of the parameters can be in the form of ones, zeros, normal distribution, truncated normal distribution, etc. A different version of gradient descent recommends is using where the difference between these versions is based on the number of data samples that feed to the network for each iteration.

- Gradient descent (GD) which is based on applying the gradient algorithm to every single observation in the training set.
- Stochastic gradient descent (SGD) is the opposite of the SG method. SGD introduces a random sample of the data on its iteration. The cons of this method are the slowness.
- Batch gradient descent is based on feeding all data to the network at the same time. The disadvantage of this method is the high risk while the positive is the processing speed.
- Mini-batch gradient descent: is based on feeding the networks with N random of a group of samples to overcome the cons on the SGD such as the acceleration processes.

During our comparisons between the optimization methods, we notice that ADAM performed better with a large amount of data while LBFGS works for smaller amounts of data since it converges faster and may perform with similar results to other stochastic optimizers.

Algorithm 2: MLP with Stochastic Gradient Descent

Input: Dataset S , feed x_i to the input layer, initialize the weights w_i , learning rate η and the bias b_i variables
Output: Compute model $y_i = f(\sum_{i=0}^k w_{ij}x_i + b_i)$

- 1 Compute the back propagation using SGD $\mathcal{L}_\epsilon(y, f(x, w)) = \frac{1}{2} \sum_{i=0}^k (y_{k_i} - \hat{y}_{k_i})^2$
- 2 **repeat**
- 3 **for** $(x, y) \in S$ **do**
- 4 $w_{i+1} = w_i - \eta \cdot \nabla J(W)$;
- 5 **for** $i \in \{1, \dots, i\} \wedge x_i \neq 0$ **do**
- 6 $w_i \leftarrow w_i - \eta (\frac{\partial}{\partial w_i} l(y(x | \Theta), y) + 2\lambda \frac{w_i}{\pi})$;
- 7 **end**
- 8 **end**
- 9 **until** stopping criterion is not met;

3.5. The Neural Network Structure

The activation function can have a general form where W is the weight vector and b is the bias variable. NNs that don't use an activation function will end up with a linear

output. The non-linear functions are preferred to learn more complicated data patterns that linear functions aren't able to do.

$$h = f(W^T X + b) \quad (26)$$

The activation function is used in the forward and backward propagation process in the network to compute the summation of the error to show the loss of the models. The number of iterations is set to 200 and the number of hidden layers and neurons in the proposed models can be shown in Table 2.

Table 2. MLP parameters.

MLP Models	Hidden Layers	Learning Rate
Model 1	3	e^{-1}
Model 2	3	e^{-2}
Model 3	5	e^{-3}
Model 4	6	e^{-4}
Model 5	6	e^{-10}
Model 6	7	e^{-20}

4. Results

In this section, the results of investigating the wireless channel modeling in the high MmWave bands are presented. We started with constructing a logarithmic path loss model using different channel state information. Furthermore, we were able to assist base stations to predict the bands and path loss using ML techniques. To elaborate on the results, we used multiple evaluation matrices to evaluate the accuracy of the ML models such as the average root mean square error (RMSE), defined as:

$$\bar{\mu} = \sqrt{\frac{1}{k} \sum_{i=1}^k \bar{\mu}_j} \quad (27)$$

$$\bar{\mu}_j = \sqrt{\frac{1}{k} \sum_{i=1}^k (y_i - h(\mathbf{x}_i, \mathbf{w}))^2} \quad (28)$$

where μ_j in the above equation can be obtained from the test set, given by and k is the number of data samples in the test set. Additionally, to evaluate the proposed models, regression evaluation metrics are performed. Mean squared error (MSE) is used to average the sum of the squared difference between the predicted label \hat{y} and the actual label y . Mean absolute error (MAE) is measured by taking the average absolute difference between the predicted label and the actual label. Low values are targeted to reach perfect prediction without errors in both MSE and MAE. R-Squared (R^2) is used to evaluate the models and to determine how close they fit the regression line by measuring the sum square error to the total square error [33]. Other statistical metrics can be obtained from the below properties

$$MSE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (29)$$

where y_i and \hat{y}_i are the actual and predicted values for multiple purposes in this manuscript such as predicting the frequency bands and the path loss of a high MmWave in macro urban environment.

4.1. Bands

With new wireless generations, the wireless bands are becoming broader, and it is essential to allow the base station to predict the bands. In this section, the base stations can predict the bands based on ML algorithms. With the use of wireless channel information, the bands can be predicted. Machine learning methods such as algorithms and deep learning techniques are implemented to infer the frequency bands. Figure 6, shows that as the bands increase, the path loss increases as well.

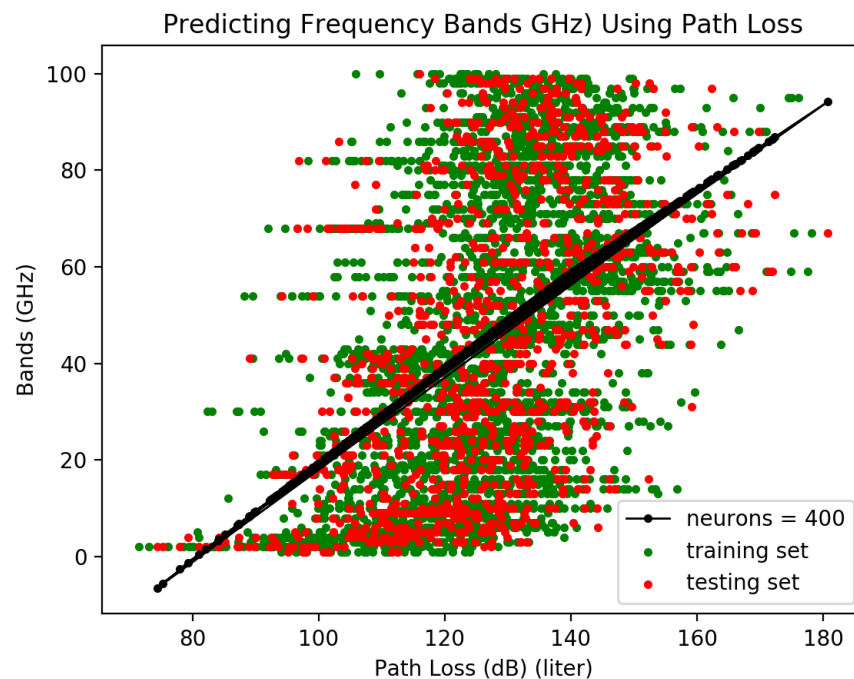


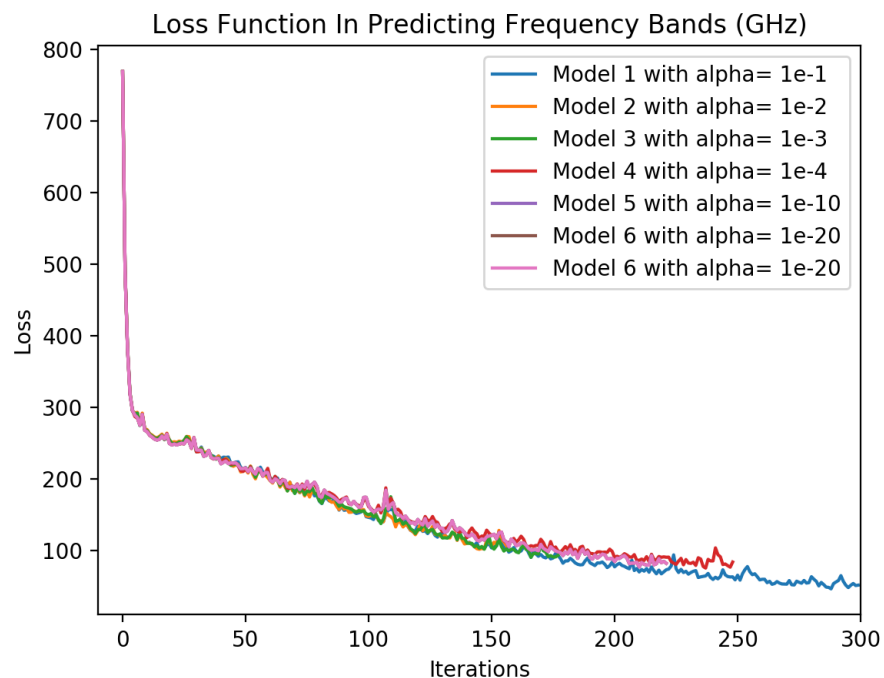
Figure 6. The relation of MmWave bands with path loss.

Figure 6 shows the prediction of frequency bands using only a CSI feature where the purpose of this figure is to show the proportion relation between path loss and frequency to meet Equation (7).

Table 3 shows the evaluations and the capability of base stations to predict the wireless frequency bands that are allocated from 1 to 100 GHz. The highest prediction was the modified schemes of the combination of RFR and PCA techniques as marginal of supervised and unsupervised learning. That combination of Random Forests regression and PCA results in a higher prediction than using the Random Forest by itself due to the capability of PCA to find the pattern of the CSI features. An illustration of how the loss function predicts the bands can be seen in Figure 7.

Table 3. Frequency bands prediction evaluation metrics.

ML Models	R-Square	MAE	MSE	RMSE
Model 1	65.6	0.73	12.2.9	17.3
Model 2	63.3	13.3	322.2	17.9
Model 3	65.0	12.9	306.8	17.5
Model 4	67.1	12.2	287.7	17.0
Model 5	65.0	12.7	307.0	17.5
Model 6	66.2	12.6	296.4	17.2
Model 7	73.0	11.7	238.9	15.45
Random Forests Regress (RFR)	83.7	10.5	142.5	11.94
RFR and PCA	90.23	9.8	132.6	3.89

**Figure 7.** The loss of the proposed models.

4.2. Path Loss

Path loss is usually obtained from the measured PDP by calculating the received power using the integration of the area under PDP. Since the transmitted power and the antenna gains are known, path loss can be obtained from the wireless channel empirically. In this manuscript, we are proposing a new methodology to allow base stations to predict the path loss based on learning using ML algorithms. In this section, investigation of the path loss in the higher frequency bands compared to the lower frequencies will be illustrated. A comparison between a 95 frequency and a lower 28 GHz frequency in similar environments will be shown in this section. Path loss is also can be predicted using NN and Random Forests and the results of that can be seen in Table 4.

Table 4. Path loss evaluation metrics.

ML Models	R-Square	MAE	MSE	RMSE
Model 1	73.63	5.94	60.81	7.79
Model 2	75.02	5.91	57.61	7.59
Model 3	73.00	6.24	62.77	7.93
Model 4	76.48	5.77	54.24	7.36
Random Forests Regress	89.18	3.34	24.93	4.99

The loss function in Equation (20) can be represented in Figure 8 that shows how the loss of predicting the path loss decayed with the number of iterations. In terms of how the loss curve is used to investigate the path loss, three models of MLP algorithms have been used, as seen below. The models are the same but with different learning rates; we can see that at a low learning rate, the improvement can be seen as linear. With a lower learning rate, the sharper the curve became, the faster the decay. A very high learning rate can cause overfitting in the model.

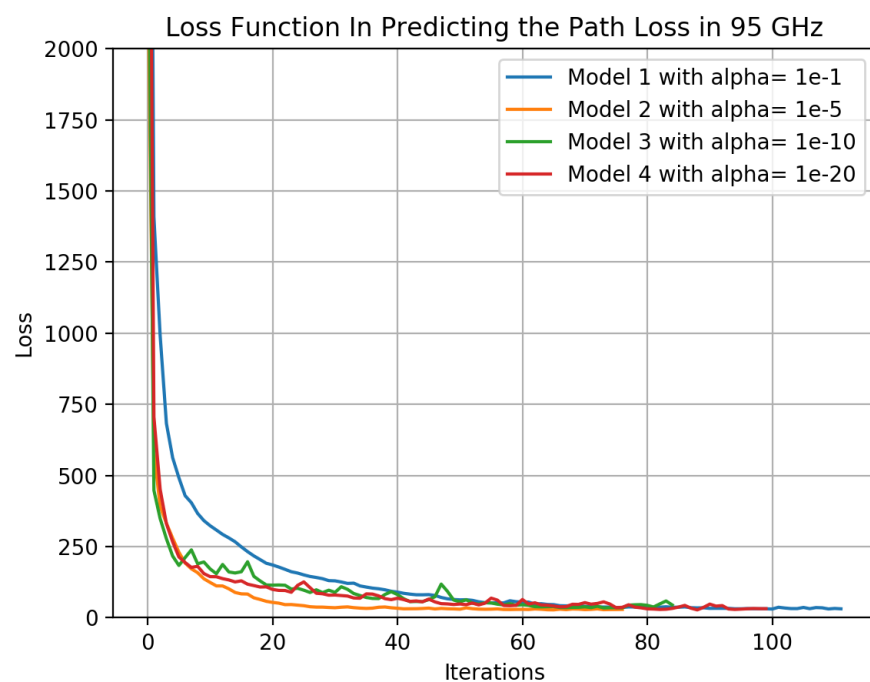
**Figure 8.** The loss of the deep learning models with different learning rate.

Table 1 contains path loss predicting evaluation in the higher mmWave bands. Random Forests obtains the highest accuracy with 90.24%, while other MLP models have lower accuracy. This is due to Random Forests using an ensemble technique that reduces the variance error. Moreover, one might notice that the normalization on that table was neglected due to the low difference between normalization process results and the regular procedure. Other features that led the Random Forests model to have the highest accuracy compared to the DL models are random subsets for each node in the decision tree and multiple features in that subset that are selected to decide the best split. Moreover, the diversity of decisions from tree estimators that are averaged to control the error made the Random Forests model the optimum one.

4.3. Validation Results

To validate the proposed methodology in order to assist the base station to be capable to predict the frequency bands, we generated another dataset. The second environment is a micro-urban environment where the generated bands were from 10 to 100 GHz with an increment of 5. The results almost matched. Table 5 shows the R-square only of both cases for the same applied models. The above procedures that were implemented on cellular systems and following the same methodology can be applied to other systems such as wireless local area networks (WLAN) and wireless sensor networks (WSN).

Table 5. Frequency bands prediction evaluation metrics using an alternative case study.

ML Models	R-Square A	R-Square B
Model 1	65.6	70.6
Model 2	63.3	70.9
Model 3	65.0	63.9
Model 4	67.1	63.8
Model 5	65.0	77.3
Model 6	66.2	76.9
Model 7	73.0	79.4
Random Forests Regress	83.7	94.9

5. Conclusions

The mmWave bands for beyond 5G have been examined in this manuscript with the methods of applying ML techniques. The superiority of ML towards wireless communication is proposed in this manuscript based on a data-driven method. By utilizing wireless channel modeling, data and other DL and ML approaches, we were able to assist the base stations in predicting the frequency bands and the path loss. The DL MLP technique was associated with the performance of other ML algorithms such as Random Forests regressor to predict path loss and frequency bands. Optimization methods were used to assess and update the internal variables such as the weight where various types of optimization algorithms can improve the system performance. The Random Forest as supervised learning was modified by an unsupervised PCA algorithm to assist the base station to enhance the prediction of the MmWave bands. Future work should indicate further comparison of other AI techniques and explore other wireless channel measurement features. Thus, based on our investigation in this manuscript, we can conclude that applying machine learning to assist the base station is a promising technique and should be implemented in current and future wireless communication systems.

Author Contributions: For research articles tasks have been distribution between authors as following: conceptualization, A.M.A.; methodology, writing and data curation, S.A.A.; supervision, K.-C.C.; review and editing, E.M.M. and visualization, A.A.-S. All authors have read and agreed to the published version of the manuscript.

Funding: Authors express a great appreciation to Prince Sattam bin Abdulaziz University, Saudi Arabia, for their support to this publication.

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

References

1. Aldossari, S. Artificial Intelligence Towards the Wireless Channel Modeling Communications in 5G. Ph.D. Thesis, University of South Florida, Tampa, FL, USA, 2020.
2. Wang, T.; Wen, C.K.; Wang, H.; Gao, F.; Jiang, T.; Jin, S. Deep Learning for Wireless Physical Layer: Opportunities and Challenges. *China Commun.* **2017**, *14*, 92–111. [[CrossRef](#)]

3. Nawaz, S.J.; Sharma, S.K.; Wyne, S.; Patwary, M.N.; Asaduzzaman, M. Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future. *IEEE Access* **2019**, *7*, 46317–46350. [[CrossRef](#)]
4. Jiang, C.; Zhang, H.; Ren, Y.; Han, Z.; Chen, K.; Hanzo, L. Machine Learning Paradigms for Next-Generation Wireless Networks. *IEEE Wirel. Commun.* **2017**, *24*, 98–105. [[CrossRef](#)]
5. Zhang, T.; Liu, S.R.; Xiang, W.; Xu, L.; Qin, K.; Yan, X.Y. A Real-Time Channel Prediction Model Based on Neural Networks for Dedicated Short-Range Communications. *Sensors* **2019**, *19*, 3541. [[CrossRef](#)] [[PubMed](#)]
6. Aldossari, S.; Chen, K.C. Relay selection for 5g new radio via artificial neural networks. In Proceedings of the 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC), Lisbon, Portugal, 24–27 November 2019; pp. 1–6.
7. Kurniawan, E.; Zhiwei, L.; Sun, S. Machine Learning-Based Channel Classification and Its Application to IEEE 802.11ad Communications. In Proceedings of the GLOBECOM 2017—2017 IEEE Global Communications Conference, Singapore, 4–8 December 2017; pp. 1–6. [[CrossRef](#)]
8. Aldossari, S.M.; Chen, K.-C. System and Method for Predicting Wireless Channel Path Loss. U.S. Patent 11,128,391, 21 September 2021.
9. 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](#)]
10. Das, S.S.; Prasad, R. *Evolution of Air Interface Towards 5G: Radio Access Technology and Performance Analysis*; River Publishers: Gistrup, Denmark, 2019.
11. Pometcu, L.; D’Errico, R. An Indoor Channel Model for High Data-Rate Communications in D-Band. *IEEE Access* **2020**, *8*, 9420–9433. [[CrossRef](#)]
12. 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](#)]
13. FC Commission. *FCC Takes Steps to Open Spectrum Horizons for New Services and Technologies*; FC Commission: Washington, DC, USA, 2019.
14. Hossain, F.; Geok, T.K.; Rahman, T.A.; Hindia, M.N.; Dimiyati, K.; Ahmed, S.; Tso, C.P.; Rahman, A.; Ziela, N. An efficient 3-D ray tracing method: Prediction of indoor radio propagation at 28 GHz in 5G network. *Electronics* **2019**, *8*, 286. [[CrossRef](#)]
15. Xing, Y.; Rappaport, T.S. Propagation Measurement System and Approach at 140 GHz-Moving to 6G and Above 100 GHz. In Proceedings of the 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 9–13 December 2018; pp. 1–6. [[CrossRef](#)]
16. Ayadi, M.; Ben Zineb, A.; Tabbane, S. A UHF Path Loss Model Using Learning Machine for Heterogeneous Networks. *IEEE Trans. Antennas Propag.* **2017**, *65*, 3675–3683. [[CrossRef](#)]
17. Liao, Y.; Shen, X.; Sun, G.; Dai, X.; Wan, S. EKF/UKF-based channel estimation for robust and reliable communications in V2V and IIoT. *EURASIP J. Wirel. Commun. Netw.* **2019**, *2019*, 144. [[CrossRef](#)]
18. Aldossari, S.; Chen, K.C. Predicting the Path Loss of Wireless Channel Models Using Machine Learning Techniques in MmWave Urban Communications. In Proceedings of the 22ND International Symposium on Wireless Personal Multimedia Communications (WPMC), Lisbon, Portugal, 24–27 November 2019; pp. 1–5.
19. Wang, J.; Jiang, C.; Zhang, H.; Ren, Y.; Chen, K.C.; Hanzo, L. Thirty years of machine learning: The road to Pareto-optimal wireless networks. *IEEE Commun. Surv. Tutor.* **2020**, *22*, 1472–1514. [[CrossRef](#)]
20. Sun, S.; Rappaport, T.S.; Thomas, T.A.; Ghosh, A.; Nguyen, H.C.; Kovács, I.Z.; Rodriguez, I.; Koymen, O.; Partyka, A. Investigation of Prediction Accuracy, Sensitivity, and Parameter Stability of Large-Scale Propagation Path Loss Models for 5G Wireless Communications. *IEEE Trans. Veh. Technol.* **2016**, *65*, 2843–2860. [[CrossRef](#)]
21. Friis, H.T. A note on a simple transmission formula. *Proc. IRE* **1946**, *34*, 254–256. [[CrossRef](#)]
22. Saunders, S.R.; Aragón-Zavala, A. *Antennas and Propagation for Wireless Communication Systems*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
23. Wang, J.; Jiang, C.; Zhang, H.; Ren, Y.; Chen, K.C.; Hanzo, L.H. Thirty Years of Machine Learning: The Road to Pareto-Optimal Next-Generation Wireless Networks. *arXiv* **2019**, arXiv:1902.01946.
24. Kibria, M.G.; Nguyen, K.; Villardi, G.P.; Zhao, O.; Ishizu, K.; Kojima, F. Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks. *IEEE Access* **2018**, *6*, 32328–32338. [[CrossRef](#)]
25. Sun, S.; Rappaport, T.S.; Rangan, S.; Thomas, T.A.; Ghosh, A.; Kovacs, I.Z.; Rodriguez, I.; Koymen, O.; Partyka, A.; Jarvelainen, J. Propagation Path Loss Models for 5G Urban Micro- and Macro-Cellular Scenarios. In Proceedings of the 2016 IEEE 83rd Vehicular Technology Conference, Nanjing, China, 15–18 May 2016; pp. 1–6. [[CrossRef](#)]
26. Dai, H.N.; Wong, R.C.W.; Wang, H.; Zheng, Z.; Vasilakos, A.V. Big Data Analytics for Large-scale Wireless Networks. *ACM Comput. Surv.* **2019**, *52*, 1–36. [[CrossRef](#)]
27. Kristem, V.; Bas, C.U.; Wang, R.; Molisch, A.F. Outdoor Wideband Channel Measurements and Modeling in the 3–18 GHz Band. *IEEE Trans. Wirel. Commun.* **2018**, *17*, 4620–4633. [[CrossRef](#)]
28. Zhang, Y.; Wen, J.; Yang, G.; He, Z.; Luo, X. Air-to-Air Path Loss Prediction Based on Machine Learning Methods in Urban Environments. *Wirel. Commun. Mob. Comput.* **2018**, *2018*, 8489326. [[CrossRef](#)]
29. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
30. Friedman, J.; Hastie, T.; Tibshirani, R. *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2001; Volume 1.

-
31. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:cs.LG/1412.6980.
 32. Fletcher, R. *Practical Methods of Optimization*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 1987.
 33. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)—Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [[CrossRef](#)]