Machine Learning-Based Path loss Models For the UAV Air-to-Air (A2A) Prediction

Ritwik NA[1], Saydul Akbar Murad[1] and Nick Rahimi[1]
[1]*School of Computing Sciences & Computer Engineering, University of Southern Mississippi, Hattiesburg, USA*
E-mail: ritwik.na@usm.edu, saydulakbar.murad@usm.edu, nick.rahimi@usm.edu,

***Abstract*—The study is centered on wireless communication, with a particular emphasis on 5G and the construction of relay systems while critically analyzing path loss. It explores the importance of path loss in trustworthy communication, taking environmental factors, frequency, and distance into account. In wireless communication, unmanned aerial vehicles (UAVs) are essential for path loss optimization in particular. The research shows advances in the use of empirical models like the Okumura-Hata model in conjunction with machine learning approaches like Random Forests and Artificial Neural Networks. To predict and comprehend path loss, the methodology blends deterministic methodologies, machine learning, and empirical methods. Various algorithms, including Random Forests, ANN, KNN, Gaussian Naive Bayes and Linear Regression, are employed, providing a comprehensive understanding of path loss modeling, addressing challenges in diverse settings, and showcasing the adaptability of machine learning in urban and UAV-assisted environments.**

Keywords : Wireless Communication, Path Loss Modeling, 5G Networks, Unmanned Aerial Vehicles (UAVs), Empirical Models, Deterministic Models, Ray-Tracing Method, Machine Learning, Urban Environments

I. Introduction

Path loss, or signal attenuation, is a critical aspect of wireless communication, describing the weakening of electromagnetic signals as they travel from a transmitter to a receiver, influenced by factors like distance, frequency, impediments, and environmental conditions. It plays a pivotal role in deploying and enhancing 5G and future relay systems. The advent of fifth-generation wireless networks (5G) emphasizes extensive coverage and reliable communication, supporting diverse Internet of Things (IoT) applications across vast areas and terrains. IoT involves interconnected physical items, sensors, and gadgets utilizing the internet for data exchange[1]. Unmanned aerial vehicles (UAVs) contribute significantly to IoT applications, with their ability to autonomously gather data, transmit information, and perform tasks. UAVs, or drones, serve as versatile platforms for path loss calculation and optimization in wireless communication, addressing various applications like traffic monitoring, logistics, and atmospheric investigations due to their cost-effectiveness and efficient wireless communication capabilities[3].

The propagation environment in UAV-assisted communication systems significantly differs from that of traditional systems, presenting a host of unique challenges. A thorough and accurate understanding of UAV wireless channels is critical to the effective design and deployment of these communication systems. In UAV communication, wireless signals to and from the UAVs can face obstructions and have varying propagation conditions throughout their transmission path[6]. Consequently, numerous research efforts have been dedicated to developing adaptable and precise path loss models tailored specifically for UAV communication scenarios. In a typical metropolitan environment, samples for various signal propagation paths are generated using sophisticated ray-tracing software[5]. This advanced software takes into careful consideration the diverse altitudes of both the transmitting (Tx) and receiving (Rx) Unmanned Aerial Vehicles (UAVs) within the urban setting. As part of this research effort, a statistical propagation model has been specifically proposed for the UAV communication channel operating at low altitudes in urban environments[5]. This model accounts for a range of critical factors, with particular attention given to the elevation angle between the UAV transmitter (Tx) soaring through the skies and the ground-based receiver (Rx). Interestingly, the research findings have revealed a crucial relationship between the accuracy of prediction results and the elevation angle, highlighting the significance of this parameter in optimizing UAV communication systems for urban applications.

II. Related Work

Table I

Summary Of Related Work

Ref.	Research Focus	Key Points
[1]	Path Loss Models in FWA	Various empirical models (SUI, ECC-33, COST-231 Hata) assessed for FWA in rural, suburban, and urban settings. The ECC-33 model demonstrated reliability and efficiency, especially in urban scenarios.
[2]	Random Forests in Path Loss Prediction	Decision trees form the basis of Random Forests, offering improved prediction accuracy. They use a random vector of values for each tree, stabilizing generalization error. The model's effectiveness is determined by the strength of individual trees, influenced by their number and size. Random Forests provide resilience by randomly selecting features for decision tree node splits.
[3]	UAVs in Wireless Communication	UAVs play a crucial role in wireless communication, offering cost-effective access and adaptability. Low-altitude UAVs have advantages in on-demand networks, demonstrating short-range line-of-sight connectivity. Understanding channel characteristics is essential for effective UAV communication systems.
[4]	Millimeter-Wave Technology for UAV Communications	Research explores millimeter-wave technology for UAV communications, addressing challenges in high-data-rate, urgent, or ad hoc scenarios. Hierarchical beamforming codebooks and spatial division multiple access enhance capacity. Adaptive UAV cruising algorithms are crucial for maintaining communication in dynamic situations.
[5]	Air-to-Ground Path Loss Model in Urban Environments	Study focuses on a statistical propagation model predicting air-to-ground path loss in urban environments for aerial wireless base stations. Different propagation groups are identified, each with a unique path loss profile. Understanding these groups is vital for optimizing air-to-ground wireless services.
[6]	Radio Channel Modeling for UAV Communication	Research aims to create models considering path loss exponents and shadowing for UAV communication within cellular networks. Simulations with a commercial UAV reveal that increasing height reduces path loss exponent, emphasizing the importance of height-dependent properties in UAV propagation channels.
[7]	Air-to-Air Channel Characteristics for UAS	Investigation into Air-to-Air channel characteristics for Unmanned Aerial Systems (UAS). Communication-aware mobility behavior proposed to address challenges. The Rice channel model is modified to consider multipath effects caused by UAV altitude, enhancing the dependability and efficiency of UAS operation.
[8]	ANN Prediction Models for Outdoor Environment	Artificial Neural Networks (ANN) used for predicting outdoor propagation path loss. Novel error correction model created by combining theoretical model with ANN. Performance evaluated using various metrics, comparing ANN models with established models like COST 231-Walfischlkegami.
[9]	LS-SVRGA Model for Radio-Wave Path Loss Prediction	Least Squares Support Vector Regression with Genetic Algorithms (LS-SVRGA) models employed for predicting radio-wave path loss in suburbia. LS-SVRGA model outperformed competitors in terms of prediction accuracy. Genetic algorithms used to optimize key LS-SVR model parameters.

III. Propagation Environment

Built-in signal propagation models can mimic multi-path propagation in a variety of settings, including open, rural, suburban, and urban areas, particularly in satellite communication scenarios[1][3]. Here's a brief explanation:

A. Satellite Elevation Zones

Open Sky Limit : Above this point, multi-path propagation is unaffected, allowing satellites to have a clear line of sight (LOS) to a ground station or receiver.

Multi-path Zone : Satellites with elevation angles between the Open Sky Limit and the Obstruction Limit are regarded as LOS signals, however multipath propagation effects are still possible. These satellites are simulated using the LOS scenarios from the ITU model[1][3].

B. Obstruction Zone

The Obstruction Zone contains satellites whose elevation angles fall below the Obstruction Limit. In this situation, direct signal pathways may be blocked, frequently by structures or the landscape. A likelihood for a non-line-of-sight (NLOS) circumstance is given to take this into consideration. Using the ITU model for NLOS circumstances, satellites categorized as NLOS based on this likelihood are simulated. It's crucial to remember that until the satellite leaves the obstruction zone, the NLOS classification remains in effect[3].

C. Elevation Mask Setting

A new restriction is added to the simulation by this setting. A satellite is taken into account in the simulation if it has a minimum elevation angle with respect to the horizon. This mask setting excludes from the simulation any satellites with elevation angles lower than this.

With the use of this method, the simulator is able to realistically simulate the effects of multi-path propagation on satellite signals under varied environmental circumstances. It makes a distinction between satellites with unobstructed LOS (above the Open Sky Limit), satellites whose LOS signals are influenced by multi-path (in the Multi-path Zone), and satellites in the Obstruction Zone, where NLOS conditions are taken into account because of probable obstacles. Based on the satellites' minimum elevation angles, the Elevation Mask setting further narrows the selection of which ones are included in the simulation.

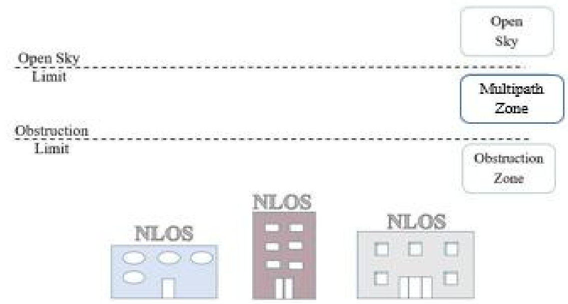


Fig. 1 : ITU Multi-path Propagation Model

IV. Methods

1. Empirical Methods:

1. A. Log-Distance Model:

The Friis free space model is extended with the log distance path loss model in order to forecast propagation loss in a wider variety of situations. The log distance model takes into consideration numerous environmental conditions, as opposed to the Friis free space model, which is restricted to open, unimpeded routes between transmitter and receiver. It is a more flexible model since it incorporates random shadowing effects brought on by obstacles like hills, trees, and buildings. It is frequently referred to as the "log normal shadowing model" since it can take into account the variability and unpredictability of signal blocking in a variety of situations[4].

$$PL(d)_{dB} = PL(d_0)_{dB} + 10_n log_{10}(\frac{d}{d_0}) + X$$

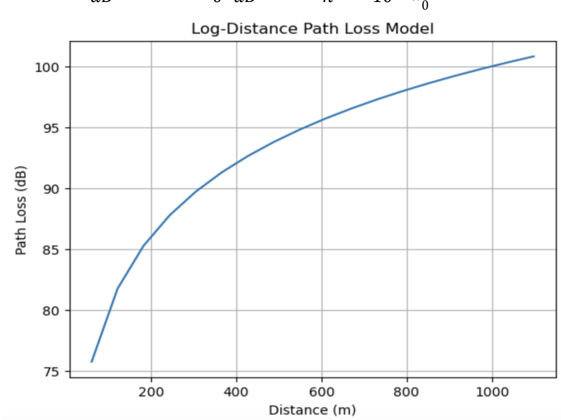


Fig. 2 : Log-Distance Path loss model

The path-loss exponent (PLE) - critical parameter in wireless communication models. Referenced PLE values may not accurately represent the specific environment. Estimating the PLE is crucial before modeling. It's determined by equating empirical and theoretical values. Neglecting shadowing effects simplifies path loss to a straight line. To incorporate shadowing, a Gaussian random variable with a zero mean and standard deviation (σ) is incorporated. Precise knowledge of PLE and σ is essential for accurate modeling due to other influencing factors.

1. B. The Okumura-Hata model:

For estimating signal propagation in metropolitan environments, the Okumura-Hata model is in fact a frequently used empirical model. It can be used for distances between one and one hundred kilometers and frequencies between 150 and 1920 MHz. It can also be extrapolated for use at up to 3 GHz. This model is useful for a variety of wireless communication applications throughout the stated frequency and distance ranges since it aids in estimating signal behavior and path loss in urban situations[1].

$$Path\ loss\ (PL) = Free\ Space\ Path\ Loss\ (FPL) + A(f, d) - G(h_t, h_r) - G\ (Area)$$

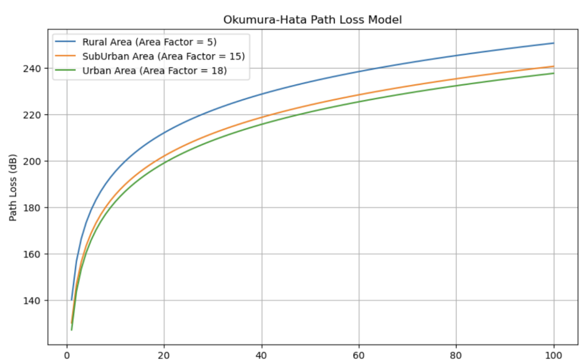


Fig. 3 : Okumura Path Loss Model For Rural, Suburban and Urban areas

Obstacles in the landscape frequently result in non-line-of-sight pathways on uneven terrain. An "**Isolated Ridge**" correction factor is used in Okumura's model to lessen the effect of particular obstacles, mainly isolated ridges. More complex terrains cannot be accurately modeled by this correction factor due to its limited use. Okumura's simple mean attenuation model cannot be easily integrated with generic models for diffraction loss calculations.

Hata-model : Based on Okumura's findings, this empirical path loss calculation is valid in the 150 MHz to 1500 MHz frequency range and is appropriate for urban settings.

$$Path\ Loss\ (dB) = 69.55 + 26.16 \times log_{10}(f) - 13.82 \cdot log_{10}(h_t) \cdot a(h_r) + (44.9 - 6.55 \cdot log_{10}(h_t) \cdot log_{10}(d)$$

For forecasting cellular transmissions in suburban and rural regions, the Okumura-Hata model—an improved variant of the Okumura Model—is extensively utilized. The addition of graphical data improves the Okumura model. With consideration for base station antenna height and operating frequency, the Hata Model forecasts total path loss for terrestrial microwave and cellular communications[4]. It is appropriate for broadcast as well as point-to-point communications.

TABLE II
ABBREVIATION TABLE FOR EMPIRICAL METHODS

Abbreviation	Meaning
$PL(d)_{dB}$	Path loss at distance d in decibels
$PL(d_0)_{dB}$	Reference path loss at reference distance d_0 in decibels
n	Path loss exponent
d	Distance at which path loss is calculated
d_0	Reference distance
X	Additional random or log-normal shadowing effects
A(f, d)	Additional path loss due to factors such as frequency (f) and distance (d)
$G(h_t, h_r)$	Gains or losses associated with transmitter height (h_t) and receiver height (h_r)
G(Area)	Gains or losses associated with specific characteristics of the area

f	Frequency in MHZ
h_t	Height of the transmitter antenna in meters
h_r	Height of the receiver antenna in meters
d	Distance between the transmitter and receiver in kilometers.
$a(h_r)$	Adjustment factor based on the height of the receiver antenna

2. Deterministic Method:

2. A. Ray - Tracing Method:

Radio channel prediction requires the use of propagation models. Both time-dispersive and non-time-dispersive deterministic models, such as ray-tracing, are possible. From channel data, time-dispersive models predict path loss. As deterministic approaches are more accurate than assumptions, they are more efficient. Three parameters are needed to estimate propagation characteristics using ray-tracing: the complex permittivity of the reflecting surface, which is a specified static quantity; the propagation distance; and the incidence angle to reflecting surfaces, which is calculated via ray-tracing[6]. The categories of ray-tracing model algorithm:

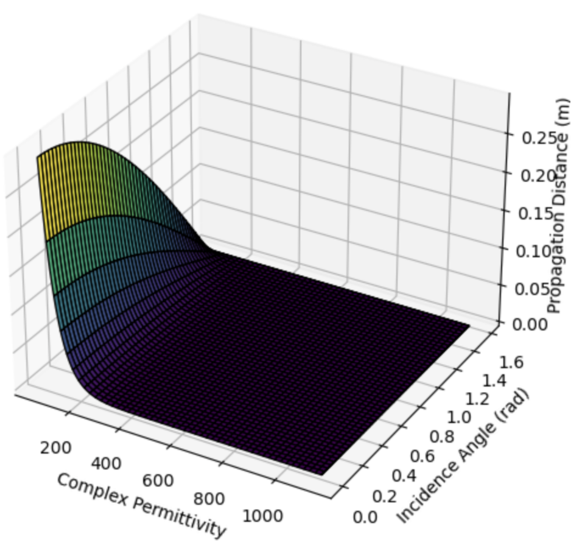


Fig. 4 : Ray-Tracing method for Radio Channel Prediction

2. A. 1. The Imaging Method:

The imaging approach combines the position of the transmitter, the position of the receiver, and the reflecting surfaces to establish a propagation path using geometric optics.

2. A. 2. Ray-Launching Method:

By launching rays at regular intervals and actively searching for the rays that eventually arrive at the receiver's site, the ray-launching method builds the propagation path.

V. Machine Learning Approach

1. Data Collection and Feature Extraction:

Samples from measurements, each with a path loss value and matching input attributes, make up the collected data. These characteristics are classified as either environment-dependent or system-dependent, depending on how they depend on the propagation environment (carrier frequency, transmitter and receiver heights and positions, etc.). The antenna separation distance and angle between the line-of-sight path and the horizontal plane are examples of further features that can be obtained from these system-dependent parameters.

The number of training samples that are available has a direct impact on the path loss model's performance. A large enough dataset should be divided into two categories as soon as possible: a training dataset for use in developing the prediction model and a test dataset for use in validating and enhancing the model.

2. Feature Selection and Scaling:

There are three main feature selection approaches: filter, wrapper, and embedding, which are related to the interaction between the feature selection process and model design[2].

2. A. Filter Approach:

This approach assesses feature relevance without taking into account the particular model. It evaluates feature importance according to predetermined standards.

2. B. Wrapper Approach:

The prediction performance is taken into account while calculating feature scores in the wrapper approach. It considers how effectively a feature adds to the predictions made by the model.

2. C. Embedded Approach:

Feature selection is easily incorporated into the modeling process using the embedded approach. It weighs the significance of each attribute while taking forecast accuracy into account.

Regression approaches, Gaussian Naive Bayes, Random Forests, Artificial Neural Networks (ANN), and K-Nearest Neighbors (KNN) are a few machine-learning algorithms that are sensitive to the size of the input field. Normalization is crucial before training to standardize input features and path loss values within a typical range, usually -1 to 1 or 0 to 1. General formulation as follows:

$$x_N = 2\left(\frac{x - x_{min}}{x_{max} - x_{min}}\right) - 1$$

3. Hyperparameter Setting and Model Training:

Predefined values known as hyperparameters are established prior to the onset of the learning process. Options include the quantity of neurons and hidden layers in artificial neural networks (ANN). To maximize the performance of path loss prediction, optimum hyperparameters must be carefully selected.

In contrast, model parameters are acquired by training samples and may vary based on the chosen learning approach. Model parameters like weights and biases are automatically learned during training.

4. Model Evaluation and Path Loss:

Using samples from the test dataset that were not used during model training, path loss models are evaluated. Evaluation measures include model complexity, generalization ability, and prediction accuracy. Common performance metrics used to quantify accuracy include correlation factor (CF), mean squared error (MSE), root mean square error (RMSE), maximum prediction error (MaxPE), mean absolute error (MAE), error standard deviation (ESD), and correlation factor (CF).

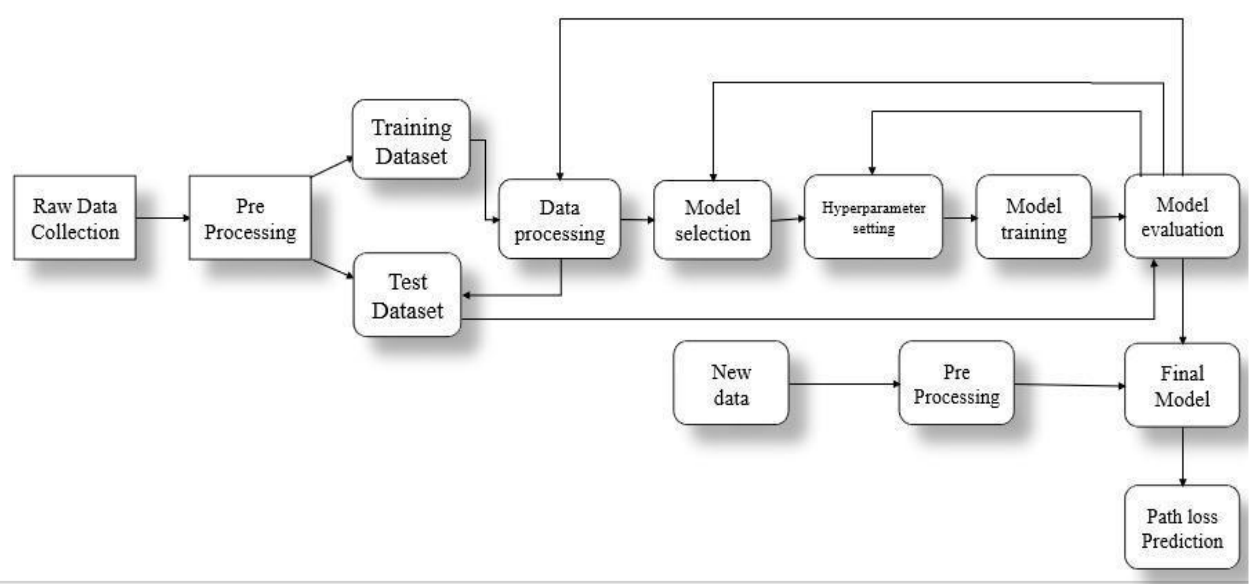
More data from various settings, such as varied terrains, frequencies, and vegetative cover conditions, can enhance the model's generalization ability.

VI. Implemented Algorithms

1. Linear Regression:

An approach for supervised learning that is commonly used for regression tasks is called linear

Fig. 5 : Machine Learning Approach



regression. Regression involves the model predicting a desired value by using independent variables. Its primary use is in forecasting and identifying correlations between variables. Regression models can change depending on the precise kind of relationship they take into account and the number of independent variables they include.

$$y = B_0 + B_1 \cdot X$$

The above formula represents a linear regression equation where "y" is the target prediction, "B_0" is the intercept and "B_1" is the coefficient for the independent variable "X".

2. K-Nearest Neighbors (KNN) :

The K-Nearest Neighbors (KNN) algorithm is a type of supervised machine learning technique that utilizes labeled input data to facilitate learning and generate suitable outputs for newly unlabeled data. Based on the basic premise that comparable things are found close together, the KNN algorithm determines that objects that are similar are located nearby to each other. Although KNN is most commonly used in the industry for classification, it may also be used for regression prediction jobs. Any technique's interpretability of output, computational efficiency, and predictive capacity are three important factors that are usually taken into account when evaluating its usefulness[9].

3. Gaussian Naive Bayes Classification :

Using Bayes' Theorem as their foundation, Gaussian Naive Bayes classifiers include a class of classification algorithms. It is actually a family of algorithms rather than a single algorithm, with all of them sharing the same guiding principle: every pair of features that is classified is treated independently of any other pair. Every characteristic contributes independently and equally to the final result, according to the basic tenet of Naive Bayes[9]. Underlying these classifiers is the Bayes Theorem, which computes the likelihood of an event based on the likelihood of an earlier occurrence. In terms of math, the expression for Bayes' Theorem is:

$$P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$

In the above formula:

- P(A|B) is the probability of event A occurring given that event B has occurred.
- P(B|A) is the probability of event B occurring given that event A has occurred.
- P(A) is the prior probability of event A.
- P(B) is the prior probability of event B.

4. Random Forest Classification :

The Random Forest Algorithm consists of multiple decision trees, each sharing identical nodes but employing distinct datasets that result in diverse leaves. By aggregating the decisions of these individual decision trees, the algorithm produces an outcome that reflects the average of the collective decisions made by all the trees.When using the Random Forest Algorithm for regression tasks, the variance or dispersion of the data within each node is measured using the mean squared error (MSE)[2].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(f_i - y_i)^2$$

In the above formula:

- n is the number of data points,
- f_i is the actual target value for the i-th data point,
- y_i is the predicted target value for the i-th data point.

5. Artificial Neural network :

An Artificial Neural Network is made up of linked nodes with activation functions arranged into layers. Among these layers is the input layer, which relays patterns to buried levels after receiving explanatory attribute values for every observation. With the potential to include one or more hidden layers, the hidden layers use weighted connections to apply changes to input values. Hidden layers send connections to the output layer, which generates response variable predictions. While hidden nodes use weighted connections to accomplish processing, input nodes are passive and duplicate values to outputs. The network differs from traditional information processing in that its ability to manipulate data effectively depends on appropriate weight selection.[8]

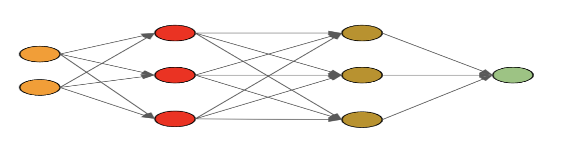
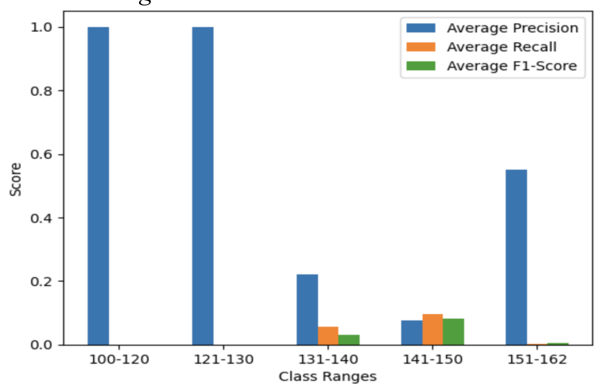


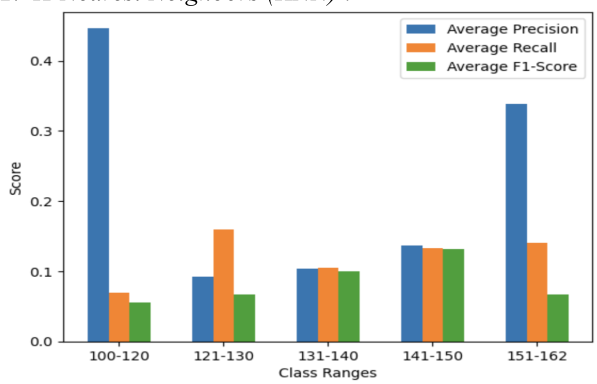
Fig. 6 : Artificial Neural Network

VII. RESULTS

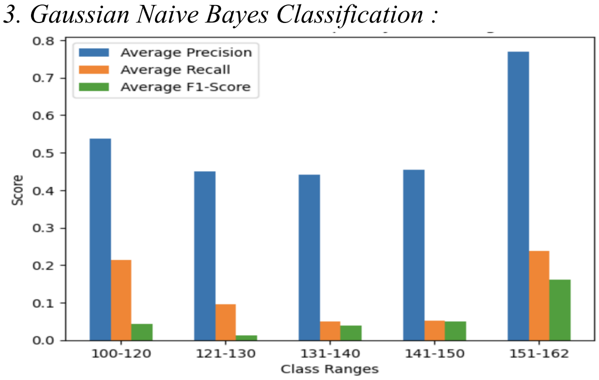
1. Linear Regression:



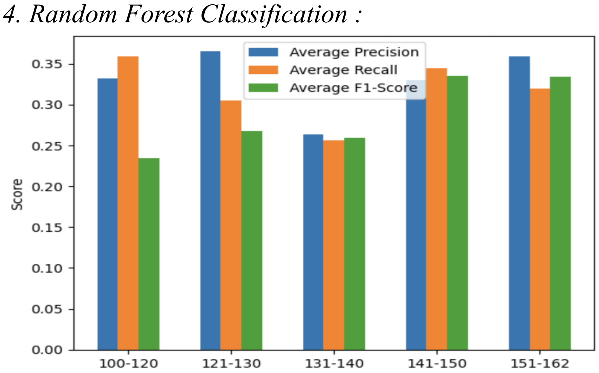
2. K-Nearest Neighbors (KNN) :



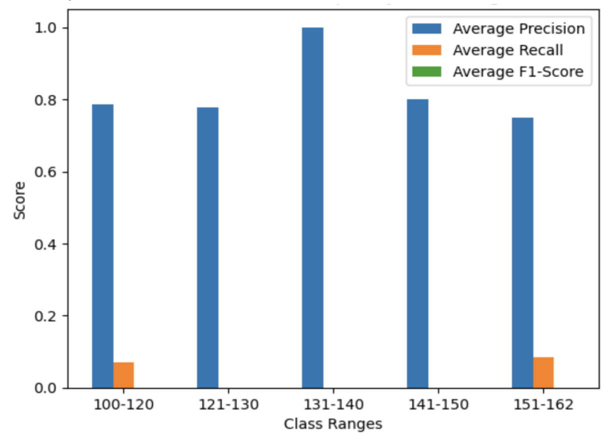
3. Gaussian Naive Bayes Classification :



4. Random Forest Classification :



5. Artificial Neural network :



VIII. CONCLUSION

The dataset contained geographical and radio propagation-related information. Each entry represents a specific location with corresponding longitude, latitude, elevation, altitude, clutter height, and distance values. Additionally, the dataset includes path loss values associated with the given locations. To understand it we researched previous works extensively and applied a broad technique that includes machine learning, empirical methodologies, and deterministic approaches. Many algorithms are used, such as Logistic Regression, KNN, Random Forests, ANN, Naive Bayes, and Linear Regression. The study offers a sophisticated understanding of path loss modeling by highlighting the applicability of machine learning techniques in urban and UAV-assisted environments. Random Forest Classifier and K-Nearest Neighbors (KNN) have the highest training accuracies (84.74% and 86.74%, respectively) among the machine learning models that have been put into practice. The efficiency of both models in identifying complex patterns in the wireless communication dataset is demonstrated by these findings. The Random Forest Classifier has good performance in identifying intricate associations, although the KNN model has a high learning potential. Based on their potential for accurate forecasts and dependable performance in real-world circumstances, these findings present both models as interesting candidates for further investigation and useful use in wireless communication path loss modeling.

IX. REFERENCES

[1] V. S. Abhayawardhana, I. J. Wassell, D. Crosby, M. P. Sellars and M. G. Brown, "Comparison of empirical propagation path loss models for fixed wireless access systems," *2005 IEEE 61st Vehicular Technology Conference*,

Stockholm, Sweden, 2005, pp. 73-77 Vol. 1, doi: 10.1109/VETECS.2005.1543252.

[2] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[3] Y. Zeng, R. Zhang and T. J. Lim, "Wireless communications with unmanned aerial vehicles: opportunities and challenges," in *IEEE Communications Magazine*, vol. 54, no. 5, pp. 36-42, May 2016, doi: 10.1109/MCOM.2016.7470933.

[4] Z. Xiao, P. Xia and X. -g. Xia, "Enabling UAV cellular with millimeter-wave communication: potentials and approaches," in IEEE Communications Magazine, vol. 54, no. 5, pp. 66-73, May 2016, doi: 10.1109/MCOM.2016.7470937.

[5] A. Al-Hourani, S. Kandeepan and A. Jamalipour, "Modeling air-to-ground path loss for low altitude platforms in urban environments," 2014 IEEE Global Communications Conference, Austin, TX, USA, 2014, pp. 2898-2904, doi: 10.1109/GLOCOM.2014.7037248.

[6] R. Amorim, H. Nguyen, P. Mogensen, I. Z. Kovács, J. Wigard and T. B. Sørensen, "Radio Channel Modeling for UAV Communication Over Cellular Networks," in IEEE Wireless Communications Letters, vol. 6, no. 4, pp. 514-517, Aug. 2017, doi: 10.1109/LWC.2017.2710045.

[7] N. Goddemeier and C. Wietfeld, "Investigation of Air-to-Air Channel Characteristics and a UAV Specific Extension to the Rice Model," 2015 IEEE Globecom Workshops (GC Wkshps), San Diego, CA, USA, 2015, pp. 1-5, doi: 10.1109/GLOCOMW.2015.7414180.

[8] I. Popescu, D. Nikitopoulos, P. Constantinou and I. Nafornita, "ANN Prediction Models for Outdoor Environment," 2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications, Helsinki, Finland, 2006, pp. 1-5, doi: 10.1109/PIMRC.2006.254270.

[9] Lin, KP., Hung, KC., Lin, JC., Wang, CK., Pai, PF. (2010). Applying Least Squares Support Vector Regression with Genetic Algorithms for Radio-Wave Path-Loss Prediction in Suburban Environment. In: Zeng, Z., Wang, J. (eds) Advances in Neural Network Research and Applications. Lecture Notes in Electrical Engineering, vol 67. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-12990-2_100