



PATH LOSS PREDICTION BASED ON MACHINE LEARNING TECHNIQUES: SUPPORT VECTOR MACHINE, ARTIFICIAL NEURAL NETWORK, AND MULTILINEAR REGRESSION MODEL

^{*1}Idogho, J. & ¹George, G.

¹Department of Computer Science, Faculty of Computing and Applied Sciences, Baze University, Jabi Abuja, Nigeria

*Corresponding Author Email: gilbert.george@bazeuniversity.edu.ng

ABSTRACT

The rapid progress in fairness, transparency, and reliability is inextricably linked to Nigeria's rise as one of the continent's leading telecom markets. Path loss has been one of the key issues in providing high-quality service in the telecommunications industry. Comparing route loss prediction systems with high accuracy and minimal complexity is so critical. In this article, the simulation of data was compared using three alternative models: Artificial Neural Network (ANN), Support Vector Machine (SVM), and a conventional Multilinear Regression (MLR) model. The performance of the various models is evaluated using measured data. The simulated outcome was then assessed using various performance efficiency metrics, including the Determination Coefficient (R^2) and Root Mean Square Error (RMSE), Mean Square Error (MSE) and Root Square Error (R^2) (MSE). For the modelling of all inputs, the anticipated results showed that the ANN model is marginally better than the SVM model. The results also demonstrated that the ANN and SVM models could model path loss prediction better than the MLR model. As a result, it is possible to recommend using ANN to estimate path loss.

Keywords: *Artificial Neural Network, Machine Learning, Multilinear Regression, Path Loss, Support Vector Machine; Wireless Sensor Network*

LICENSE: This work by Open Journals Nigeria is licensed and published under the Creative Commons Attribution License 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided this article is duly cited.

COPYRIGHT: The Author(s) completely retain the copyright of this published article.

OPEN ACCESS: The Author(s) approves that this article remains permanently online in the open access (OA) model.

QA: This Article is published in line with "COPE (Committee on Publication Ethics) and PIE (Publication Integrity & Ethics)".

INTRODUCTION

Path loss is a sort of signal blurring that occurs when the signal intensity between the receiving and transmitting stations is reduced. GSM, like other wireless communication systems, relies on radio waves propagating through the lower atmosphere. In different places, the transmission path between a transmitter and a mobile receiver varies, ranging from a simple view path to one with obstacles such as slopes, trees, structures, and other man-made structures. Propagation loss owing to reflection, refraction, diffraction, ingestion, and scattering affects the electric field intensity of signals emerging from a transmitter in mobile communication, weak received signals and route loss due to a drop in the power density of an electromagnetic wave as it transmits from the broadcasting antenna to the receiving antenna is a major concern (Cavalcanti, et al., 2012). The population of Nigeria was estimated to be around 209,646,216 people in February 2021. (Worldometers.info, 2021). According to the Nigerian Communications Commission (ncc.gov.ng, 2021), Nigeria has grown to become the largest telecoms market in Africa and the Middle East, with more than 198 million active endusers in July 2020. The rapid progress in fairness, transparency, and reliability is inextricably linked to Nigeria's rise as one of the continent's leading telecom markets. The country's diverse telecommunications client base has risen over 1000 percent from 2003 to a current level of 187,275,547 subscribers (ncc.gov.ng, 2021). Nigeria now has four GSM operators: MTN, Globacom, Airtel, and Etisalat. With over 80 million customers, MTN values the greatest assistance. 2021) (Akinyoade *et al.*, 2017) Nigeria's GSM industry was Africa's fastest-growing telecoms market between 2003 and 2006. The telecommunications market is becoming increasingly competitive as suppliers compete for comparable potential end-users. Twenty (20) years after the GSM period began in Nigeria, the center is gradually shifting from providing inclusiveness to providing excellent services. The thrill of owning a telephone set is gradually giving way to complaints about dropped calls and jams.

Scientists and architects have spent decades developing signal propagation models that can predict signal route loss under various scenarios. There are three types of broadcast models: empirical, deterministic, and semi-deterministic. Empirical models, such as millimetre-wave propagation models and classical models, rely on estimates to show the relationship between path loss and climate parameters in a quantitative way. The deterministic models rely on mathematical theories to determine the field intensity of a vast area of rays at receiving focus, specifically, instantaneous, reflected, diffracted, and scattered rays. The two-beam model, which simulates both direct and ground reflected rays, is the most basic deterministic model. Semi-deterministic models combine the advantages of empirical and deterministic models. When compared to deterministic models, empirical models are easier to implement and require less computing power; nevertheless, empirical models are naturally touchy and less precise, whereas deterministic models require more computing power and climatic data to conduct accurate path loss expectations (Chen et al., 2017). Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are artificial intelligence techniques that are useful in the construction of models for solving prediction issues. Sadrmomtazi et al. (2013) found that ANN is flexible and can learn the underlying links between a process's inputs and outputs without needing explicit

knowledge of how these variables are related (Abiodun et al. 2017). As a result, it can be beneficial in the construction of path loss models.

STATEMENT OF THE PROBLEM, AIM AND OBJECTIVES

Propagation path loss models can forecast conventional signal intensity at a certain distance from the transmitter, estimate radio coverage regions of Base Transceiver Stations (BTS), allocate frequencies, perform interference analysis, optimize handovers, and modify power levels. Due to changes in environmental structures, local topographical profiles, and weather conditions, the path loss prediction model for a specific environment utilizing any of the existing basic empirical models has been shown to differ from the machine learning approach applicable to such an environment. (Abiodun and colleagues, 2017). As a result, it's critical to develop a more accurate model for path loss forecasting utilizing proper machine learning methods. The paper aims to compare the following machine learning procedures (**artificial neural network, support vector machine, and multilinear regression**) to discover which predicts path loss best.

OBJECTIVES

- To compare the machine learning algorithms listed below (artificial neural network, support vector machine and multilinear regression).
- To compare the performance of several machine learning models.
- Determination Coefficient (R^2) and Root Mean Square Error (RMSE), Root Square (R^2), and Mean Square Error (MSE).

RELATED LITERATURE REVIEW

Machine Learning-Based Path Loss Prediction

The path loss model converts input features into output values (path loss observations). This is critical to figure out how to derive a more generalized model that changes depending on conditional inputs in addition to developing a predictor that can make accurate predictions. While the ANN-MLP-based nonlinear model concentrates on predicting the path loss value with accuracy, the principal component analysis (PCA) and variance analysis balance it out to generate a more generalist model.

Data Pre-processing

The training data that is given to an ANN determines how accurate it will be. In addition to the model's processing and tuning capabilities, obtaining an accurate model necessitates a well-distributed, ample, and carefully measured collection of data. In light of this, preparing data is an essential step in developing an ANN learning model. Sampling and normalization are also carried out to save time and eliminate bias. Finding the ideal weights for a set of learning data is the aim of learning to produce precise predictions. Normalizing the magnitude of the input numbers is essential for obtaining the right weight because it lessens scale-related negative effects. For instance, a variable magnitude of inputs with 0.001 and 0.1 can provide a fairly large gradient, with results of 0.5 and 0.005 and a climb of 0.0005. If the input parameters are not adequately normalized, backpropagation utilizing iterative partial derivatives throughout

MLP-NN might lead to biased weights. Based on the input features' propagation characteristics, we adjusted the frequency (MHz) and distance (m) data logarithmically to balance their various scales.

After dividing the data into train and test sets, sampling bias is compensated for using cross-validation rounds. A resampling approach called cross-validation is used to assess machine learning models on a small sample of data. Only one parameter, k , determines how many groups the given data sample is divided into in the process. K -fold cross-validation is a common name for the procedure as a result. To prepare learning data, uniform random sampling is utilized to divide all measured data into two sets: training (80%) and testing (20%). The test set is used to change hyperparameters for model optimization. Model for Path Loss (2.3) A weighted network of latent variables is learned using backpropagation by the ANN, a nonlinear regression system. The ANN model handles more dimensions than the look-up table method and outperforms the regression analysis model in terms of prediction performance (Ojo et al., 2022). When taking into account the intricate propagation due to variable heights and the complex distribution of structures in metropolitan settings, the nonlinear model can match with linear regression better. We employed the logistic sigmoid function as an activation function in ANN networks.

Shadowing Model

When there are obstacles between the transmitter and receiver, such as those caused by scattering, reflection, diffraction, and absorption, the signal strength is reduced. Over long distances, path loss affects receiver power, whereas shadowing affects receiver power as a result of the formation of an obstruction or the length of it. In the past, shadowing was thought to be a normal distribution coefficient because barriers' characteristics are hard to generalize and their effects on path loss are minimal. On the other hand, shadowing effects are crucial when analyzing path loss prediction with a specific confidence level in outdoor settings with a variety of obstructions.

RELATED WORKS

Ojo *et al.*, 2022 proposed ensemble approaches for path loss estimates based on machine learning. In particular, ensemble approaches have been established to enhance the performance and accuracy of signal prediction. More network parameters were implemented and enhanced in the input layers of the multilayer perceptron neural network and radial basis function models, respectively. The bagging ensemble path loss prediction model that was created had the fewest errors across all datasets. The bagging ensemble approach, which projected path loss closest to observed data and is suitable for nearly precise path loss predictions, functions as a variance and error reduction mechanism. With respect to the observations, the created bagging ensemble path loss prediction model produced the lowest errors (MSE = 0.0011 dB, SSE = 0.6069 dB, MAE = 0.0245 dB, & R = 0.7484 dB). The bagging ensemble approach, which projected path loss closest to observed data and is suitable for nearly precise path loss predictions, functions as a variance and error reduction mechanism. Ojo *et al.*, 2021 suggest using machine learning methods for path loss estimates. First, experimental data were gathered using drive tests in six base transceiver stations for multi-transmitter situations, and the path loss of the received signal level was calculated and examined. The multilayer perceptron neural network and the radial basis function neural network are the developed path loss prediction models. Additionally, the measured path loss was used to compare the MLPNN and RBFNN models, and the results showed that the RBFNN

has a lower root mean squared errors than the MLPNN, suggesting that it was more accurate. The RBFNN shows the most accurate results.

The implementation of a machine learning-based method for path loss prediction for massive intelligent surface-assisted wireless communication in a smart radio environment is presented in this research. Using the training dataset, the path loss prediction models are developed using two bagging ensemble approaches, K-nearest neighbour and random forest (Elshennawy, 2022). A route loss model is built due to the similarities between the big intelligent surface-assisted wireless communication and the reflector antenna system to create the sequence data without having to undertake measurement campaigns. The reflector antenna system's system gain is utilized to calculate a straightforward path loss equation, which is then used to produce data samples. To confirm the prediction accuracy of the path loss prediction models, simulation results are shown. The R^2 score, mean absolute error, and root mean square error are some of the complexity and accuracy metrics used to evaluate the prediction abilities of trained route loss models. The ability of machine learning-based models to deliver excellent prediction accuracy and tolerable complexity is demonstrated. In comparison to the random forest technique, the K-nearest neighbour approach performs better and has lower prediction errors.

Surajudeen-Bakinde *et al.*, 2018 proved that route loss prediction is an important factor in radio organization planning and development since it helps to understand how radio waves behave in a specific propagation environment. A balance between ease of use and accuracy is required when using the few models that are currently available for path loss estimates. This study demonstrated the development of a new route loss prediction model based on an Adaptive Neuro-Fuzzy Inference System (ANFIS) that is suitable for Very High Frequency (VHF) groups and multi-transmitter radio propagation scenarios. To determine the power advantages of radio signals received from three different transmitters, field estimations were made throughout three testing driving routes in the city of Ilorin, Kwara State, Nigeria. With individual Root Mean Square Error (RMSE), Standard Deviation Error (SDE), and connection coefficient (R) upsides of 4.45 dB, 4.47 dB, and 0.92, the developed ANFIS-based path loss model produced a low prediction error. The ANFIS-based model showed good generalization ability when it was used to predict path loss in a different but comparable propagation situation, with RMSE, SDE, and R upsides of 4.46 dB, 4.49 dB, and 0.91, respectively. Overall, the proposed ANFIS-based path loss model provides advantages in terms of simplicity, high prediction precision, and strong generalization capacity—all essential features for evaluating radio inclusion and interference feasibility during multi-transmitter radio organization arranging in the VHF groups.

Wang & Lei (2022) constructed a wireless channel path loss model based on LS-SVM and offer a simulated annealing approach to optimize the kernel function and regularization function parameters in LS-SVM. The SA+LS-SVM model is demonstrably superior to the pre-optimized model and the traditional model in prediction accuracy and can more accurately predict the fading change to the channel condition, according to the results of a MATLAB simulation experiment comparing the optimized and improved model based on the SA algorithm with the traditional model.

MATERIALS AND METHODS

The investigation was conducted at BAZE University, which has a population of around 7,000 people (7,000). Blocks of lightly constructed structures with estimated average heights of five (5) to ten (10) meters make up the environment. In addition, the environment is partially surrounded by running water and averagely raised tree vegetation. Around 60% of the land is occupied by structures made of concrete blocks, tiles, and bricks.

EXPERIMENTAL SETUP

A site investigation exercise was done using testing tools. The measurement tools used are:

1. A TEMS (Transmission Evaluation and Monitoring System) phone from Ericsson.
2. The GPS (Global Positioning System) (GPS).
3. TEMS software loaded on a laptop.

To get the network's received signal strength level, Ericsson Transmission Evaluation and Monitoring System (TEMS) phones with a sensitivity of 110dBm were utilized to initiate calls.

A location tracking device is the Global Positioning System (GPS). It uses a Global Positioning Satellite to track its whereabouts. Longitude and latitude are displayed on the GPS gadget in different formats. It calculates the distance between a location and a reference location as well as the velocity of the body on which it is placed. The GPS was employed to track the location of the study site for this experiment. It is also employed

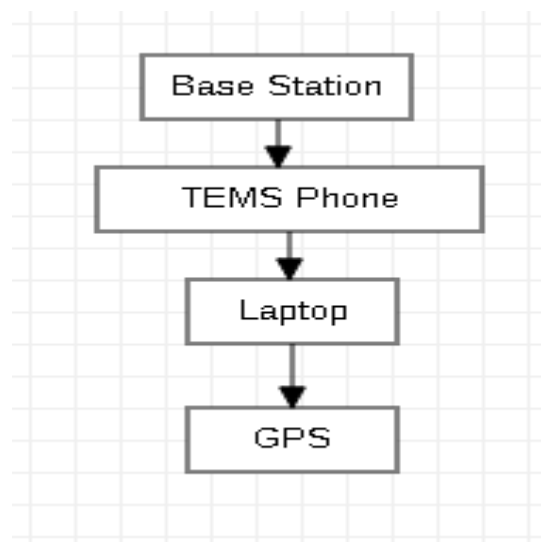


Figure 1: Shows the block diagram of the measurement procedures and experimental setup respectively

THEORY OF MODELS WITH EQUATION AND ARCHITECTURE

Machine learning is the most fast-moving innovation lately. Machine learning is a part of artificial intelligence (AI) and computer science that centres on the utilization of data and algorithms to emulate the way that people learn, step by step improving their precision (IBM Cloud Education, 2020).

For this research, the following models will be used for modelling the data.

1. Artificial Neural Network (ANN)
2. Support Vector Machine (SVM)
3. Multi-linear regression (MLR)

Artificial Neural Network (ANN)

You can think of artificial neural networks as networks of "neurons" with hierarchical algorithms. The lower layer is made up of predictors (or inputs), and the higher layer is made up of predictions (or outputs). A middle layer that has "hidden neurons" may also exist. Additionally, ANN has shown effectiveness in a variety of disciplines when dealing with complex functions. Among them are control energy classification, prediction, organization, forecasting, and simulation. (Miau, 2017.) Artificial Neural networks can be thought of as networks of "neurons" with hierarchical algorithms. Predictors (or inputs) make up the lower layer and predictions (or outputs) make up the upper layer. There may also be a middle layer that contains "hidden neurons." Additionally, ANN has shown to be efficient in working with complex functions in different fields. Some of these include classification, prediction, arrangement, forecasting, and simulation of control energy. (Miau, 2017.)

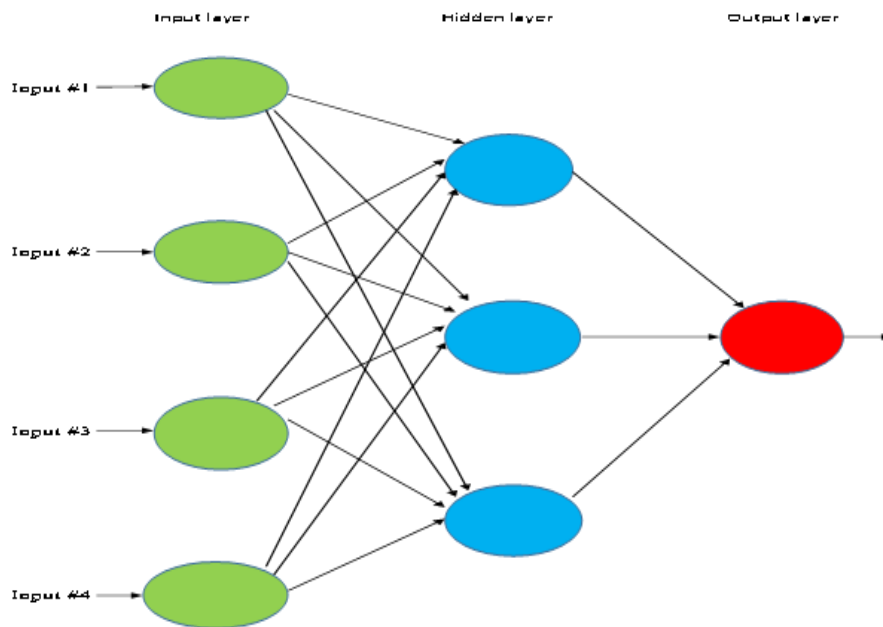


Figure 2: Feed forward ANN

$$j = b_j + \sum_{i=1}^4 W_{i,j} X_i \quad (1)$$

$$S(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

The data is used to "learn" the parameters $w_1, 1, \dots, w_4, 3$ and $b_1, b_2,$ and b_3 . To keep the weights from growing out of control, their values are frequently constrained. It is common practice to set the weight restriction setting to 0.1. (Miau, 2017) The weights take random values, to begin with, and these are then updated using the observed data. Consequently, there is an element of randomness in the predictions produced by a neural network. Therefore, the network is usually trained several times using different random starting points, and the results are averaged. The number of hidden layers was given as $(n + 1) = 5$ and must be specified in advance. The cross-validation was placed as all data was computed using 75% for training and 25% for testing.

Support Vector Machine

A type of ML technique that relies on quantifiable learning hypotheses is called a support vector machine (SVM). The fundamental idea behind SVM is to nonlinearly map a large amount of data from a low-dimensional space to a high-dimensional space to the point where the dataset is distinguishable. SVM is intended to handle relapse difficulties as an extension, therefore it might be used for way misfortune expectation (Yan et al., 2019).

Multi-linear regression

Another name for the statistical method used to predict the result of a variable based on the values of two or more variables is multi-linear regression. It is a development of linear regression and is occasionally just referred to as multiple regression. The factors used to predict the value of the dependent variable are known as independent or explanatory variables, whilst the variable to be predicted is known as the dependent variable. On the other hand, a "multilinear" model is one in which Y (i.e., the outcome) is a vector of different outcomes rather than a single number. These models would be suitable for modelling a group of inputs into several outcomes. As reported by (Miau, 2017), the MRL formula is given as:

$$Y_i = B_0 + B_{1x_{i1}} + B_{2x_{i2}} + \dots + B_{px_{ip}} + \epsilon \quad (3)$$

where:

Y_i independent or predicted variable

B_0 is the y-intercept, i.e., the value of y when both x_1 and x_2 are 0.

B_1 and B_2 are the regression coefficients representing the change in y relative to a one-unit change in x_1 and x_2 , respectively.

B_p is the slope coefficient for each independent variable

ϵ is the model's random error (residual) term.

The reason for selecting the algorithms

The main reason for the selection of the ML algorithms is according to (Miau, n.d.; Ojo et al., 2022; Wang & Lei, 2022) they have performed well in different terrains.

DATA COLLECTION

All necessary setups on the Transmission Evaluation and Monitoring System (TEMS) equipment were completed before the driving test began. The signal strength information transmitted through the air interface between the base station and the mobile station was read and recorded in a log file in all three sectors by initiating calls at each test point at a reference distance of 50m until it was established. Additionally, GPS data was recorded, making it simple to compute the radial distance between the base station and the mobile station, as well as the location coordinates and altitude. Received signal strength (RSS) was measured at a reference distance of 50m from the base station and subsequent intervals for each sector were tested. The experimental data were collected at distances ranging from 50 to 800 meters in general, as the measurement differs at the same distance in each of the monitored sectors. The log files obtained from the experiment are shown in Figures 2 to 4.

The corresponding path loss in dB from the received signal strength (dBm) is calculated using equation (4)

$$PL = 20\log(d) + 20\log(f) + 32.44 - BSTx - MSRx \quad (4)$$

Where PL is the path loss in dB at a distance.

F = Frequency

D = Distance (m)

BS Tx = Base Station transmission power

MS Rx = Mobile Station receiver power

The field measurements from the base station transmitter were carried out along three different routes, designated as radio paths a, b and c as depicted in table 3.1

Table 3: Field measurements from sectors A, B and C

The numerical analysis of path loss using the path loss prediction methods is presented. Given:

f = transmission frequency = 900MHz

h_b = base station antenna height = 100m,

h_m = mobile station antenna height = 2m,

d = base station to mobile separation distance = 100m – 800m,

G_b = base station antenna gain = 1,

G_m = mobile antenna gain = 1.

DATA STANDARDIZATION

Typically, the data used for machine learning has many features. Both leaving out relevant data and keeping irrelevant features might lead to poor predictor quality. The purpose of feature selection is to choose the best subset with the fewest number of characteristics that contribute the most to learning accuracy.

There are four (4) feature selection methodologies used in this work for the feature selection process and model design, including statistical (mean, standard deviation, etc.), correlation, normalization, and visualization. When evaluating feature importance, the statistical approach is independent of the proposed model. For every data, the mean, standard deviation, skewness, and kurtosis were determined.

Some machine-learning algorithms, such as ANN, SVR, and others, are sensitive to the input space size. As a result, the normalization procedure was completed before the start of the training. That is, all input characteristics and path loss values were adjusted to fall between 0 and 1. Figure 5 demonstrates that several of the inputs are zeros after normalization. As a result, four and two inputs were computed and analysed, respectively. The normalization method chosen during this experiment is shown in equation (2) below:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (5)$$

were.

X_i = number to be normalized

X_{min} = the minimum number

X_{max} = maximum number

Table 2: Normalized value of all data

D (m)	BS TX power (dBm)	MS Rx power (dBm)	Frequency (Hz)	Path Loss (PL) (dBm)
0	0	0.8	0	0
0.06667	0	0.73333	0	0.16632
0.13333	0	0.8	0	0.19751
0.2	0	0.4	0	0.49896
0.26667	0	0.53333	0	0.45738
0.33333	0	0.2	0	0.69855
0.4	0	0.5	0	0.53846
0.46667	0	0.7	0	0.43867
0.53333	0	0.8	0	0.39709

PERFORMANCE EVALUATION METHOD FOR MODELS

Samples from the test dataset that were not used in the model training are used to evaluate the performance of machine-learning-based path loss models. Among the evaluation, metrics are prediction accuracy, generalization ability, and complexity. To evaluate the models' accuracy, performance metrics including mean absolute error (MAE), R squared (R^2), root mean square error (RMSE), and R error (R) were utilized. Computational complexity is frequently evaluated using processing time and memory usage. For instance, crucial factors that affect ANN processing time are the number of iterations and convergence speed during the training phase.

We can choose the machine learning algorithm, modify the hyperparameters, and enhance the prediction model based on the findings. Path loss values can be generated with fresh inputs after the optimal model has been built. R-squared (R^2) is a statistical measure that quantifies the amount of variation explained by an independent variable or variables in a regression model for a dependent variable.

$$R^2 = \frac{\sum (\sigma - p)^2}{(\sigma - \bar{o})^2} \quad (6)$$

Error in the Root Mean Squared (RMSE) Taking the root of mean squared error yields root mean squared error. In other terms, it is the residuals' standard deviation. RMSE is comparable to MAE, except it magnifies and punishes significant errors more harshly.

$$RMSE = \sqrt{\frac{1}{N} \sum_{j=1}^N (o - p)^2} \quad (7)$$

Mean Squared Error (MSE): measure the average squared difference between the estimated values and the actual value.

$$MSE = \frac{1}{N} \sum_{j=1}^N (o - p)^2 \quad (8)$$

RESULTS AND DISCUSSION

The best model architecture for ANN and SVM was optimized and selected utilizing a trial-and-error method based on hyper-tuning factors such as hidden neurons, activation function, learning algorithms, and so on. The requirements of most statistical evaluation criteria are met by a competent model (Abdullahi et al. 2020). The model simulation was evaluated using the most used performance metrics, such as R^2 , MSE, RMSE, and R, during both calibration and verification. It's worth noting that the simulation for all the model development was done in MATLAB 9.3. (R2020a). Meanwhile, based on the aforesaid model combination, MLR models were created in Excel software (M1 and M2). In terms of evaluative assessment, the simulated results are based on model combinations (M1 and M2) for ANN seen in Figures 6 and 7 respectively and SVM models, respectively.

Table 1: ANN-M1&M2 Result

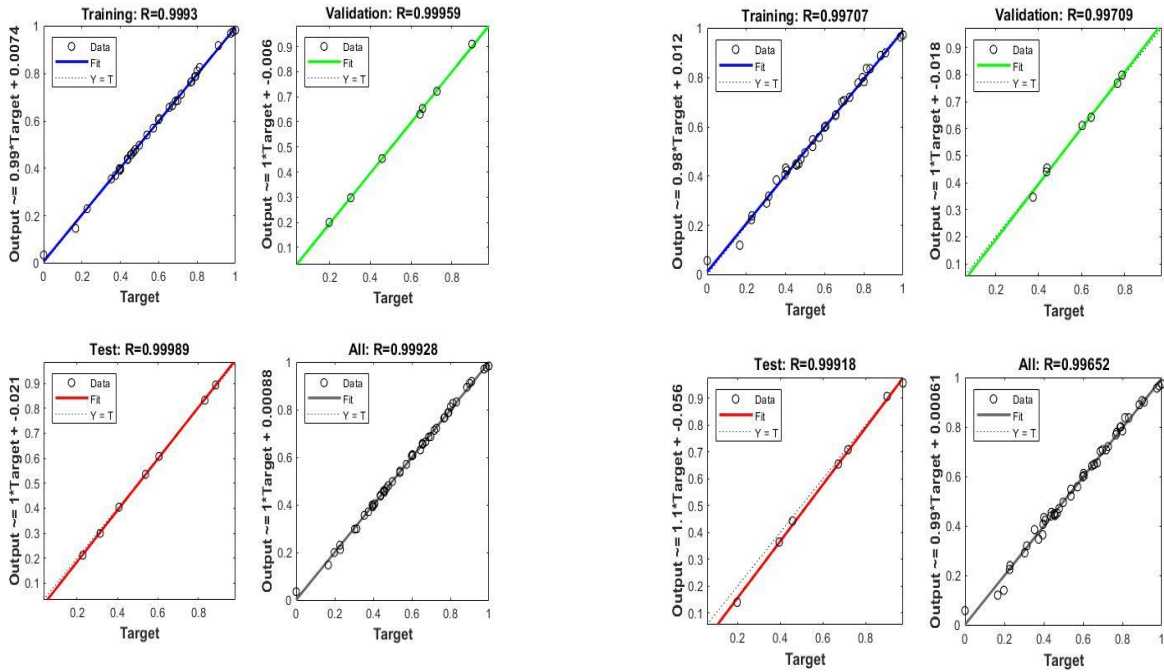
D (m)	BS TX power (dBm)	MS Rx power (dBm)	Frequency (Hz)	Path Loss (PL) (dBm)	ANN-M2 (four inputs)	ANN - M1 (two inputs)
0	0	0.8	0	0	0.05276	0.02591
0.06667	0	0.73333	0	0.16632	0.1562	0.1519
0.13333	0	0.8	0	0.19751	0.16659	0.20602
0.2	0	0.4	0	0.49896	0.4888	0.50198
0.26667	0	0.53333	0	0.45738	0.44168	0.45866
0.33333	0	0.2	0	0.69855	0.68893	0.69319
0.4	0	0.5	0	0.53846	0.53131	0.53368
0.46667	0	0.7	0	0.43867	0.42851	0.43708
0.53333	0	0.8	0	0.39709	0.39935	0.3933

Table 2: SVM-M1&M2 Result

T-R (m)	BS TX power (dBm)	MS Rx power (dBm)	Frequency (Hz)	Path Loss (PL) (dBm)	SMV - M2 (four inputs)	SMV - M1 (two inputs)
0	0	0.8	0	0	0.02758	0.02764
0.06667	0	0.73333	0	0.16632	0.16303	0.15903
0.13333	0	0.8	0	0.19751	0.18842	0.18874
0.2	0	0.4	0	0.49896	0.47895	0.4782
0.26667	0	0.53333	0	0.45738	0.4585	0.45438
0.33333	0	0.2	0	0.69855	0.67456	0.67591
0.4	0	0.5	0	0.53846	0.5446	0.54233
0.46667	0	0.7	0	0.43867	0.46581	0.46283
0.53333	0	0.8	0	0.39709	0.42255	0.42436

Table 3: MLR-M1&M2 Result

T-R (m)	BS TX power (dBm)	MS Rx power (dBm)	Frequency (Hz)	Path Loss (PL) (dBm)	MLR-M2 (Four inputs)	MLR-M1 (Two inputs)
0	0	0.8	0	0	0	0.13767
0.06667	0	0.73333	0	0.16632	0.22092	0.20693
0.13333	0	0.8	0	0.19751	0.20769	0.19391
0.2	0	0.4	0	0.49896	0.47644	0.46883
0.26667	0	0.53333	0	0.45738	0.42291	0.41468
0.33333	0	0.2	0	0.69855	0.65138	0.64847
0.4	0	0.5	0	0.53846	0.49715	0.49149
0.46667	0	0.7	0	0.43867	0.40334	0.39621
0.53333	0	0.8	0	0.39709	0.36996	0.36263
0.6	0	0.76667	0	0.43659	0.41715	0.41131
0.66667	0	0.76667	0	0.45322	0.44419	0.43943
0.73333	0	0.86667	0	0.40748	0.41081	0.40585



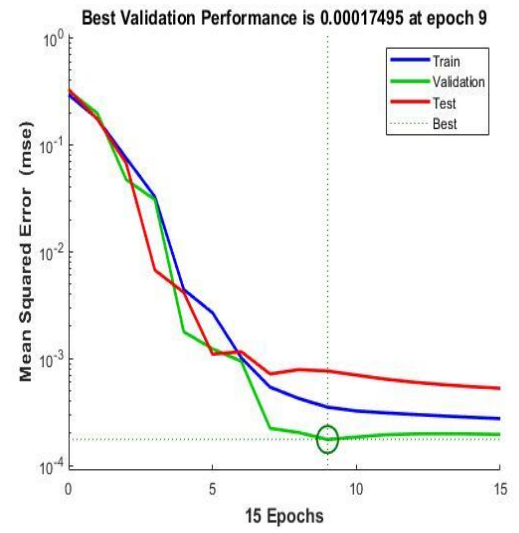
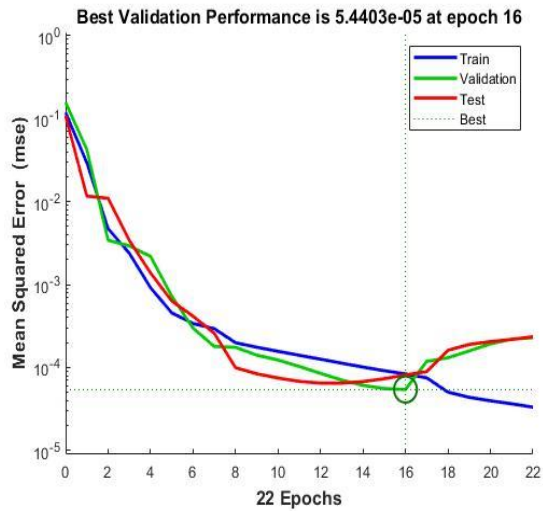


Figure 3: Showing the simulated pattern of the training function of ANN model.

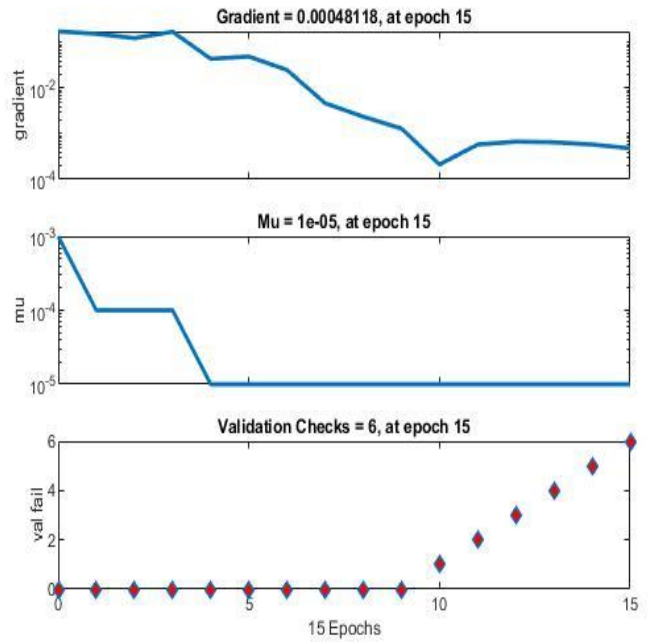
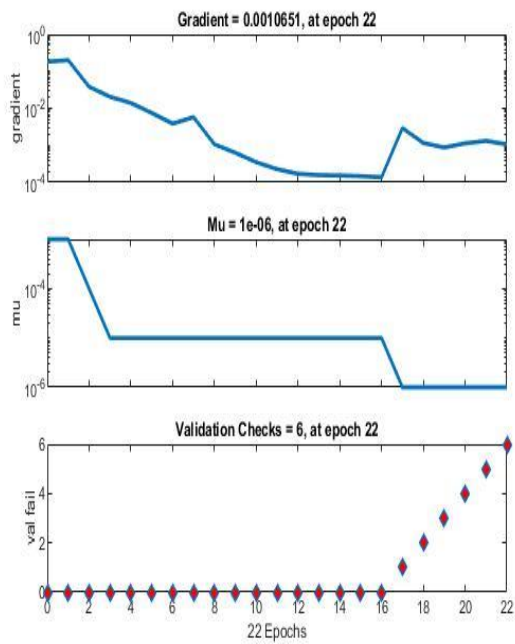


Figure 4: Training functions of stimulation of ANN (two and four inputs)

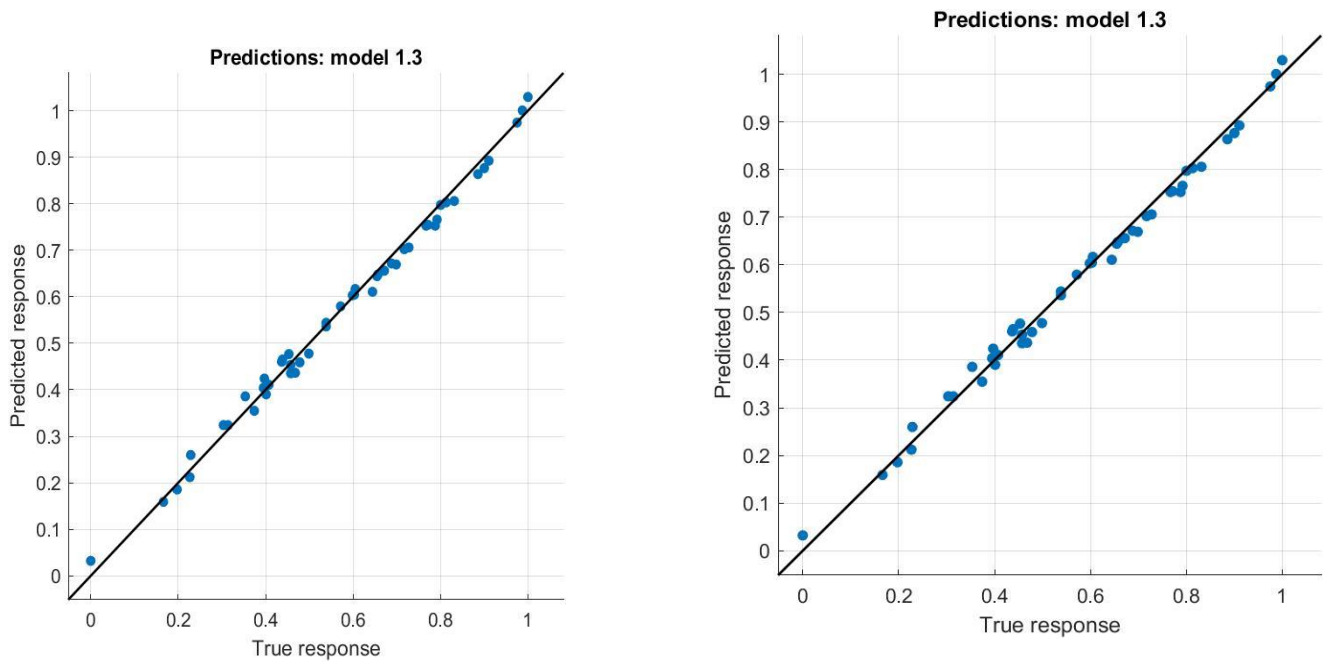


Figure 5: SVM training and testing plot (two and four inputs)

The model appears to have used the Trainlm - Levenberg–Marquardt method for training, as seen in figure 5. With varying era numbers, learned work was used as the adaptability learning capacity. The results of the two models' correlation showed that the ANN model is superior and more efficient for recreating and measuring the absorbance. This conclusion was reached by considering the R^2 and RMSE values during the preparation and testing stages. The ANN model's precognitive abilities may be related to its ability to deal with complex and exceptionally complex cycles. As a result, this approach was the driving force behind ANN's superior prediction capabilities over SVM and MLR. Abdullahi *et al.* (2020).

Comparison Analysis Using the Error Evaluation Methods

Despite the non-linear relationship between indicators (input components) and their related objective targets, the overall precision of ANN-M1 was satisfactory (low error values) and exhibited substantial conviction in terms of MLR and SVM. It's worth noting that the favourable and promising results occurred during the verification step, which is often used to precisely alter models based on known data components and objectives. The testing cycle, on the other hand, is critical in evaluating a model's exhibition since it verifies the model's accuracy based on its objective qualities.

Table 6: Predictive ability of both the linear and non-linear models

Training	Testing				Testing			
	R ²	R	MSE	RMSE	R ²	R	MSE	RMSE
ANN-M1	0.998816	0.999408	7.12082E-05	0.008438	0.99935	0.999675	2.43E-05	0.00493
ANN-M2	0.989232	0.994602	0.00064734	0.025443	0.997014	0.998506	1.12E-04	0.010568
SVM-M1	0.993907	0.996949	0.000366295	0.019139	0.993612	0.996801	2.39E-04	0.015459
SVM-M2	0.993051	0.99652	0.000417738	0.020439	0.994307	0.997149	2.13E-04	0.014593
MLR-M1	0.955911	0.977707	0.002650586	0.051484	0.966205	0.982958	1.26E-03	0.035555
MLR-M2	0.963528	0.981595	0.002192647	0.046826	0.976587	0.988224	8.76E-04	0.029595

Table 6 shows the prediction ability of both linear and non-linear models for the simulation of absorbance based on various models. The sum of squared errors is frequently stated as the cost function to be minimized by altering model parameters in the data assimilation area, and RMSE satisfies the triangle inequality that is required for a distance function metric for model evaluation. The relative mean square value avoids error compensation by indicating the on

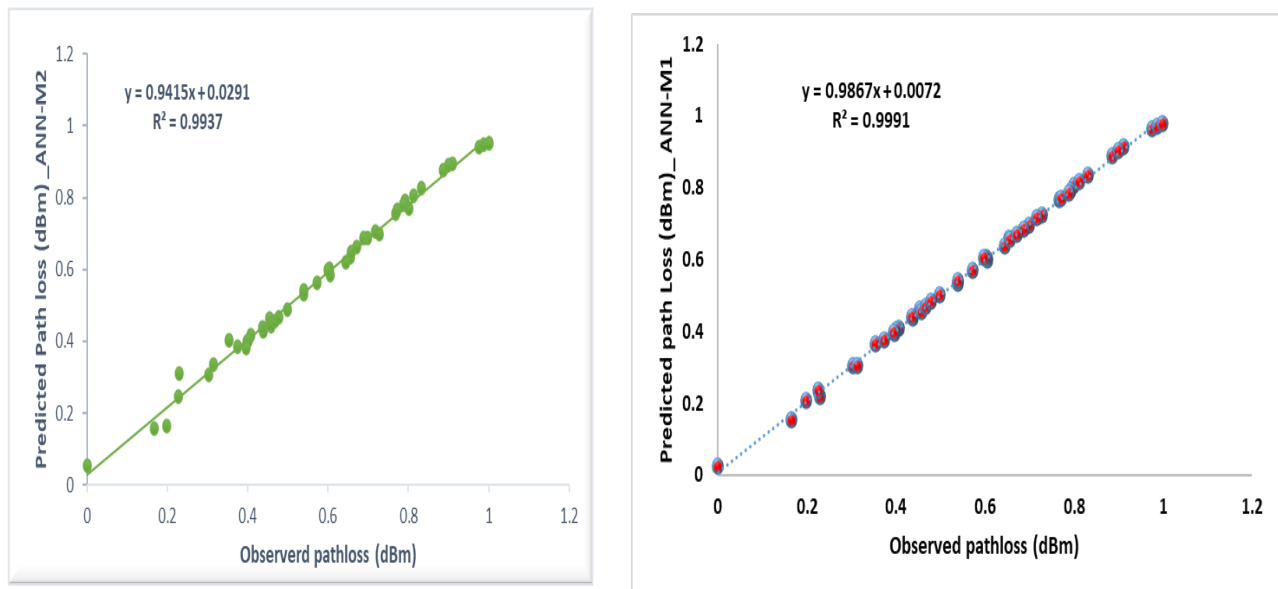


Figure 6: Scatter plot for ANN (two and four inputs)

the other hand, demonstrated the ability of both linear mathematical and nonlinear artificial intelligence models to simulate data in terms of predicting skill.

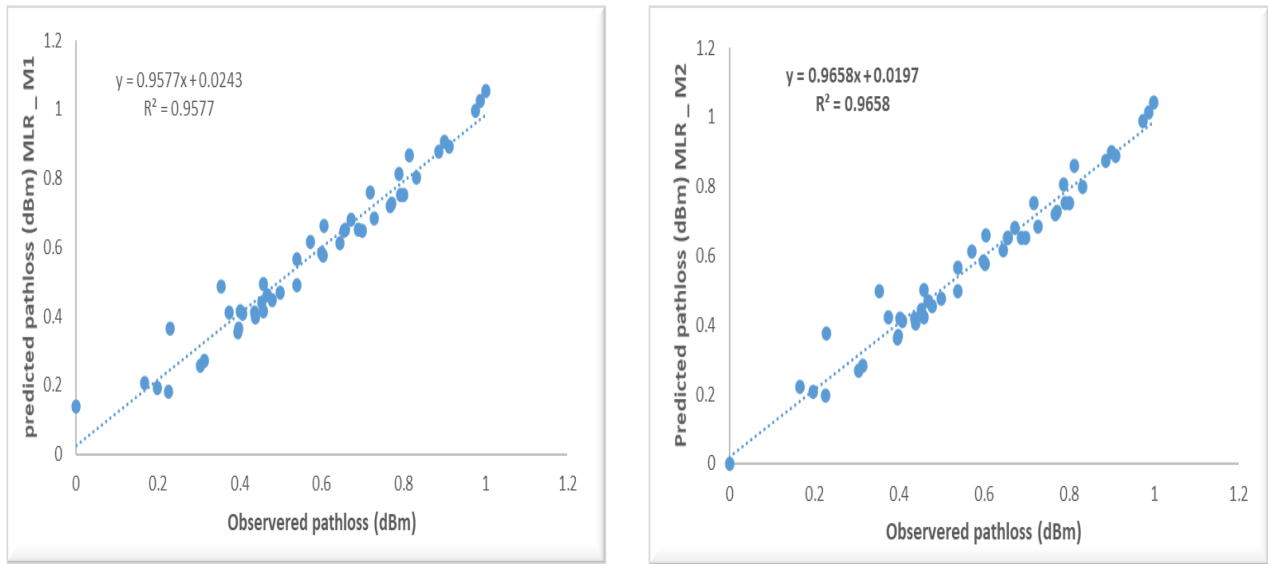


Figure 7: Scatter plot for SVM (two and four inputs)

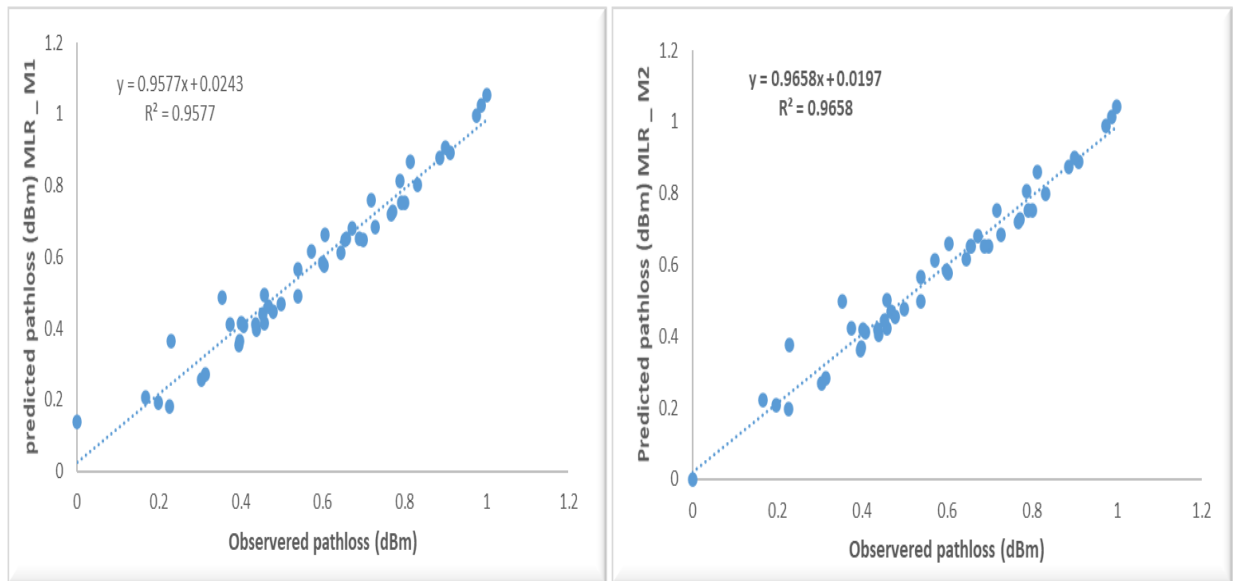


Figure 8: Scatter plot for MLR (two inputs and four inputs)

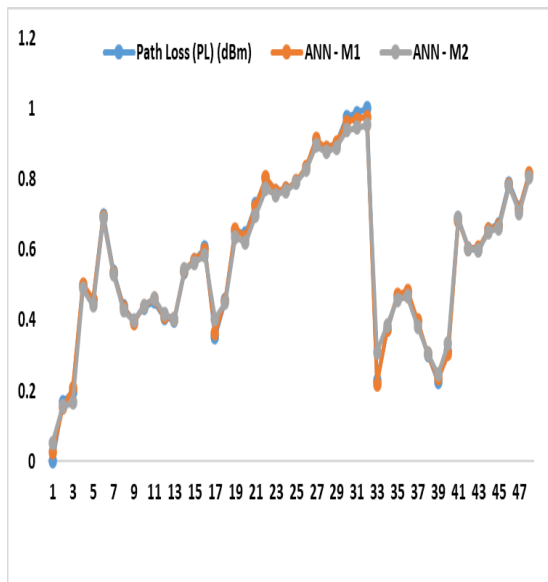


Figure 9: Time series plot for ANN models

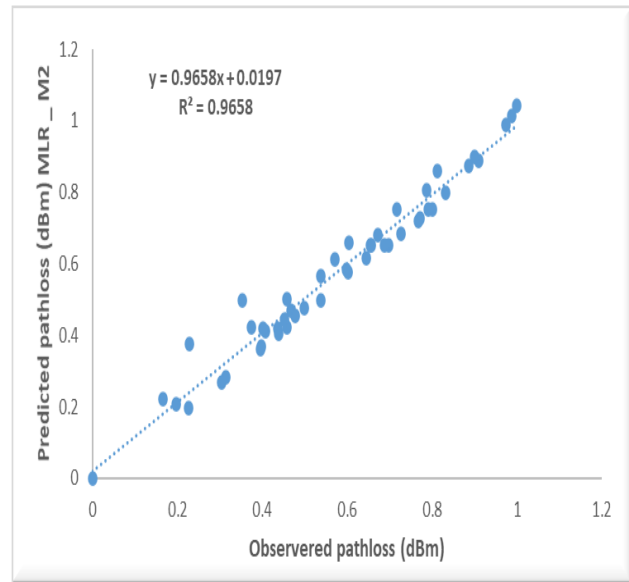


Figure 10: Time series plot for SMV models

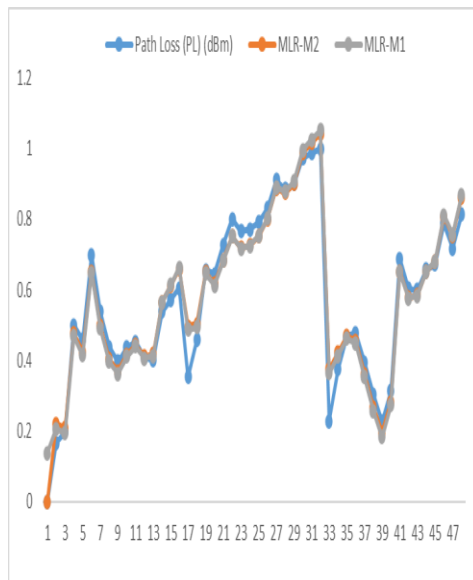


Figure 11. Time series plot for MLR models

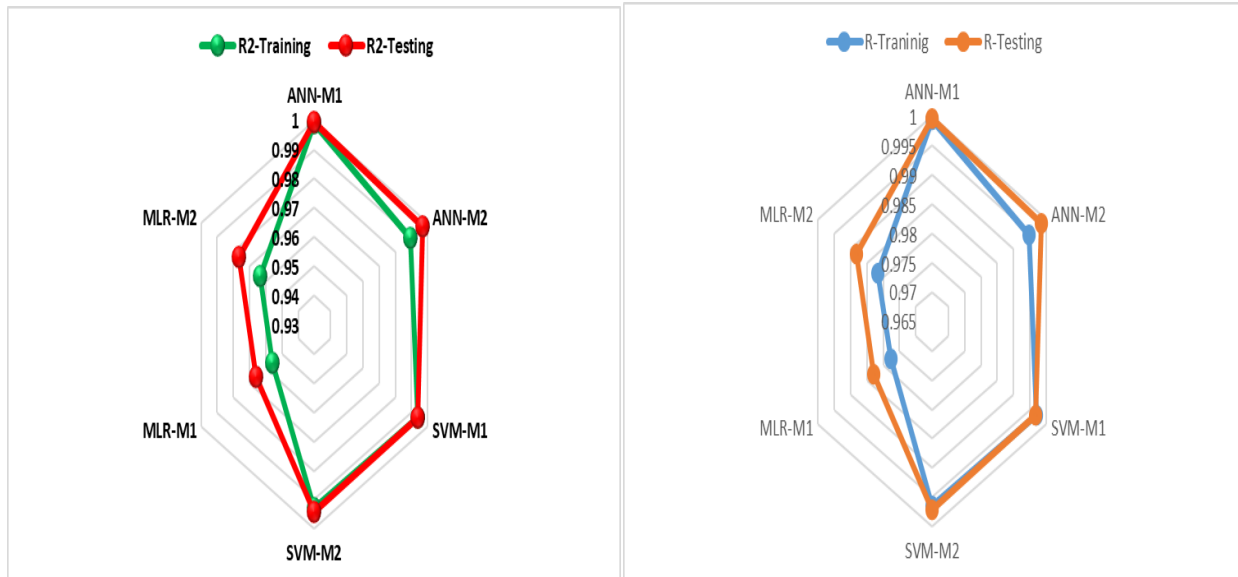


Figure 12: Radar chart showing the R² and R levels for ANN, SVM, and MLR models

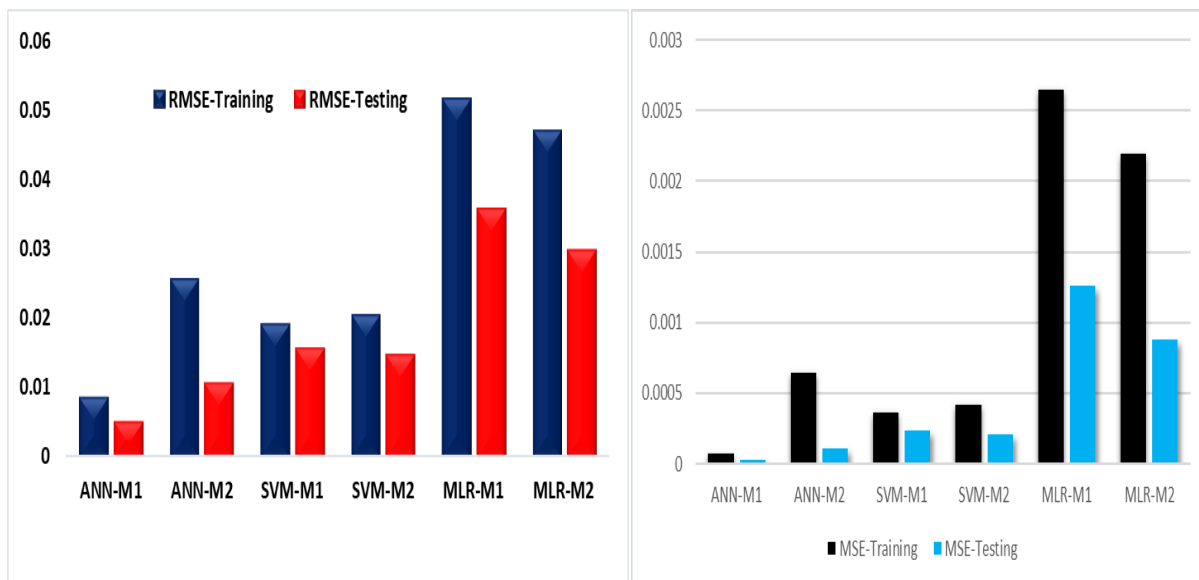


Figure 13: Error Plots in terms of MSE and RMSE

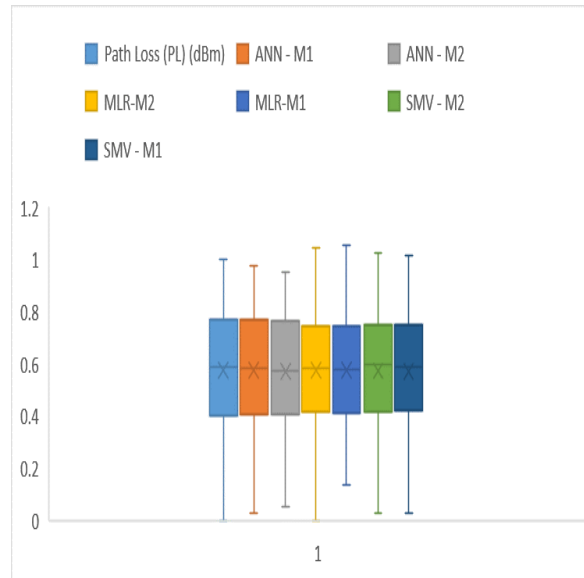


Figure 14: Box plot for all error evaluation

The ANN model assessment using primary request measurements, the correlation factor, training, and testing shows that, for the given drive test information gathered in Baze University, utilizing the given info boundaries, the ANN models perform very well in comparison to other comparative intricacy models, such as SVM and MRL models, as shown in fig 6,7 and 8.

For the simulation of all three sectors, the obtained results showed that the ANN model slightly outperformed the SVM and MLR models. The predictive findings also showed that the two models can model prediction based on the model's performance efficiency.

CONCLUSION AND RECOMMENDATIONS

This research at Baze University shows the use of three AI-based path loss prediction algorithms. In terms of performance index and correlation with actual path loss values, ML approaches such as ANN and SVM beat the MLR model, according to the analysis. The findings show that ANN is more effective at forecasting route loss. ML-based plans also do not require mathematical restrictions or unique site data to anticipate path loss, unlike observational and deterministic models. Then, in signal strength testing, allowing the application of machine learning for dependable and precise path loss estimation. Given the findings of this study, it is suggested that machine learning be used to estimate path loss because it is more effective and requires less data input. Robust ways for better tuning and selection of hyper-parameters resulting in optimal performance of ML-based systems are needed as future improvements to this research. It is possible to investigate PL prediction via incremental training, in which additional training data becomes accessible over time. Under the heading of Collection of Training Data, it has been highlighted that obtaining suitable preliminary information is critical for the accuracy and speculation of the AI-based model.

Having looked at the previous works done by (Ojo et al., 2022), and the results from our work, we can say that using deep learning models is highly recommended for future predictions of path loss in open areas and urban centers, hence

recommended. For future work, we recommend more experiments should be conducted from various terrains using deep learning algorithms.

FUNDING

This study was sponsored by the monetary contributions of the authors.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

- Abiodun, C. , Azi, S. , Ojo, J. , Akinyemi, P. (2017). 'Assessment of Path Loss Prediction Models for Wireless Propagation Channels at L-Band Frequency over Different Micro-Cellular Environments of Ekiti State, Southwestern Nigeria'. World Academy of Science, Engineering and Technology, Open Science Index 130, International Journal of Electronics and Communication Engineering, 11(10), 1103 - 1109.
- Cavalcanti, B. J., Cavalcante, G. A., Mendonca, L. M., Cantanhede, G. M., Oliveira, M. M., & Adaildo G., D. (2012). A hybrid path loss prediction model based on artificial neural networks using empirical models for LTE and LTE-A at 800MHz and 2600 MHz. *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*. From <https://doi.org/10.1590/2179-10742017v16i3925>
- Akinyoade, A., Dietz, T., & Uche, C. U. (Eds.). (2017). *Entrepreneurship in Africa*. Brill. doi:10.1163/9789004351615
- Elshennawy, W. (2022). Large Intelligent Surface-Assisted Wireless Communication and Path Loss Prediction Model Based on Electromagnetics and Machine Learning Algorithms. *Progress In Electromagnetics Research C*, **119**: 65–79. <https://doi.org/10.2528/PIERC22013002>
- IBM Cloud Education. (2020, July 2020). Retrieved June 2021, 2021 from IBM: <https://www.ibm.com/cloud/learn/machine-learning>
- Miau, S. A. (2017). New Sales Forecasting Model for International Restaurants.
- Ojo, S, Akkaya, M, Sopuru, JC. An ensemble machine learning approach for enhanced path loss predictions for 4G LTE wireless networks. *Int J Commun Syst*. 2022; 35(7):e5101. doi:10.1002/dac.5101
- Ojo, S, Imoize, A, Alienyi, D. (2021). Radial basis function neural network path loss prediction model for LTE networks in multi-transmitter signal propagation environments. *International Journal of Communication Systems*, **34**:3. <https://doi.org/10.1002/dac.4680>
- .Surajudeen-Bakinde, N. T., Faruk, N., Popoola, S. I., Salman, M. A., Oloyede, A. A., Olawoyin, L. A., & Calafate, C. T. (2018). Path loss predictions for multi-transmitter radio propagation in VHF bands using Adaptive Neuro-Fuzzy Inference System. *Engineering Science and Technology, an International Journal*, **21**(4): 679–691.
- Wang, A., & Lei, S. (2022). Research on Modeling of Underground Path Loss Prediction Based on SA-SVM. *Journal of Physics: Conference Series*, **2245**(1), 012016. <https://doi.org/10.1088/1742-6596/2245/1/012016>

Wang, R., Lu, J., Xu, Y., Shen, D., Chen, G., Pham, K., & Blasch, E. (2018, May). Intelligent path loss prediction engine design using machine learning in the urban outdoor environment. In *Sensors and Systems for Space Applications XI* (Vol. 10641, pp. 160-166). SPIE.