Progress In Electromagnetics Research C, Vol. 146, 1-12, 2024

PIER C

(Received 9 June 2024, Accepted 19 July 2024, Scheduled 30 July 2024)

Enhanced Prediction of Metamaterial Antenna Parameters Using Advanced Machine Learning Regression Models

Prince Jain[1], Prabodh K. Sahoo[1], Aymen D. Khaleel[2], and Ahmed Jamal Abdullah Al-Gburi[3,*]

[1] *Department of Mechatronics Engineering, Parul Institute of Technology Parul University, Vadodara 391760, Gujarat, India*
[2] *Computer Engineering Department, College of Engineering, Al-Iraqia University, Iraq*
[3] *Center for Telecommunication Research & Innovation (CeTRI) Faculty of Electronics and Computer Technology and Engineering Universiti Teknikal Malaysia Melaka (UTeM) Jalan Hang Tuah Jaya, Durian Tunggal 76100, Melaka, Malaysia*

ABSTRACT: The integration of machine learning (ML) regression models in predicting the parameters of metamaterial antennas significantly reduces the design time required for optimizing antenna performance compared to traditional simulation tools. Metamaterial antennas, known for overcoming the bandwidth constraints of small antennas, benefit greatly from these advanced predictive models. This study applies and evaluates four ML regression models — Extra Trees, Random Forest, XGBoost, and CatBoost — to predict key antenna parameters such as S_{11}, gain, and bandwidth. Each model's performance is assessed using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), R-squared (R^2), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE) across different training and testing set configurations (30%, 50%, and 70%). The Extra Trees model achieves the best performance for predicting gain, with an R^2 of 0.9990, MAE of 0.0069, MSE of 0.0002, RMSE of 0.0145, and MAPE of 0.3106. Feature importance analysis reveals that specific features, such as pr and $p0$ for gain and Ya and Xa for bandwidth, are critical in the predictive models. These findings highlight the potential of ML methods to improve the efficiency and accuracy of metamaterial antenna design.

1. INTRODUCTION

As wireless communication systems continue to advance, the demand for low-profile and compact antennas with high gain and wide frequency bands grows. Microstrip patch antennas have emerged as a solution, offering compactness, low profile, and light weight. Despite their advantages, these designs suffer from low efficiency, low gain, and narrow bandwidth, which recent developments in machine learning (ML) are addressing [1, 2]. Researchers are investigating methods to enhance bandwidth by modifying the dimensions and configuration of the radiating patch and decreasing the substrate's dielectric constant. Integrating notches and slots into the patch design can also improve radiation and overall performance [3]. The bandwidth of patch antennas is closely related to their size, posing a significant challenge: maintaining performance while reducing size. Researchers have attempted to keep the permittivity of the antenna high while decreasing size through the use of dielectric substrates. Additionally, modifying the antenna's shape to extend the electrical path length of the patch can increase bandwidth [4]. Adjusting the location of notches and utilizing slits and slots further reduces antenna size and improves performance.

The advancement of antennas through computational electromagnetics (EM) can be improved by utilizing ML methods to harness their inherent nonlinearities. ML's strength lies in its ability to extract meaningful information from data, making it closely related to statistics and data science. ML can build autonomous systems that match and exceed human capabilities, driven by data. However, the absence of a standardized dataset with antenna design parameters presents a challenge for machine learning engineers. Simulation software can be used to generate extensive datasets for training ML models. Designing EM devices, such as microstrip antennas, presents two primary challenges: the complexity of multi-physics analysis and the significant computational resources required. Finite element analysis (FEA) is often used but requires significant computational resources and time. ML models can predict design parameters more efficiently, especially non-parametric models, which are better suited for the highly nonlinear nature of electromagnetic devices [5].

Metamaterials, with their unique properties not found in natural materials, are widely used in various fields, including antenna design [6, 7]. These materials enable the creation of antennas with innovative features, such as increased bandwidth and gain, reduced electrical size, and improved signal directionality. Metamaterials can serve as substrates or be integrated into antenna designs to enhance performance. Simulation software, such as HFSS and CST Microwave Studio, are typically used to estimate metamaterial effects, but this process can be time-consuming. ML models can expedite the design process, reducing errors, maintaining high accuracy, saving time, and

* Corresponding author: Ahmed Jamal Abdullah Al-Gburi (ahmedjamal@ieee.org, engahmed_jamall@yahoo.com).

TABLE 1. Descriptive statistics of features in the metamaterial antenna dataset, including minimum, maximum, mean, and standard deviation values for each feature.

	Feature	Description	Maximum Value	Minimum Value	Mean	Standard Deviation
1	Wm	SRR width and height	6964.3	2142.9	2227.49	632.98
2	$W0m$	Gap between rings	651.43	162.86	401.43	184.53
3	Dm	Distance between rings	488.57	77.143	275.64	150.77
4	Tm	Width of rings	696.43	214.29	222.75	63.30
5	SRR_num	Number of SRR cells	7	3	4.10	1.44
6	Xa	Distance between antenna patch and array	10776.0	0.0	4077.50	3281.90
7	Ya	Distance between SRR cells	16607.0	2142.9	6964.33	5132.79
8	rows	Number of SRR cells in a array	7	3	4.10	1.44
9	pr	Power radiated by antenna	0.22953846	0.05481	0.19342	0.04613
10	$p0$	Power accepted by antenna	0.49982594	0.18629	0.45745	0.08890
11	Gain	Gain	3.2385	1.1987	2.7075	0.4765
12	Bandwidth	Bandwidth	124.7401	62.16	118.3604	10.34
13	S_{11}	Return loss	-2.0834	-33.9031	-16.15	7.87

improving computational efficiency. Numerous studies have validated the potential of ML models in various applications. For instance, Rawal et al. highlighted the efficacy of ML models, particularly the k-nearest neighbor algorithm, in optimizing the performance of flexible photodetectors [8]. Watpade et al. demonstrated the use of ML models, such as random forest, extra trees, decision trees, gradient boosting, and XGBoost, in predicting the dielectric constants of nano epoxy composites [9]. Jain et al. showed the effectiveness of ML models in predicting the performance features of EDM of Ti-6Al-4 V Grade-5 alloy [10]. Other studies have confirmed the accuracy and efficiency of ML models, such as XGBoost, in predicting the dielectric properties of epoxy polymer composites with various nano fillers [11–13].

This study investigates the application of four popular machine learning regression models — Extra Trees, Random Forest, XGBoost, and CatBoost — to predict critical antenna parameters, namely S_{11}, gain, and bandwidth. By evaluating these models under different data splits (30%, 50%, and 70% for testing), the aim is to determine the most efficient model and configuration for accurate and reliable antenna parameter prediction. The evaluation of each model's performance is conducted using several metrics, such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), R-squared (R^2), and mean absolute percentage error (MAPE). The results demonstrate that the Extra Trees model achieves the highest predictive accuracy, with R^2 values of 0.9984 for S_{11}, 0.9990 for gain, and 0.9859 for bandwidth. Additionally, the analysis of feature importance underscores the significant impact of certain features, such as pr and $p0$, in influencing the predictive accuracy of these models. This study underscores the potential of ML to enhance the efficiency and accuracy of metamaterial antenna design, offering a robust alternative to traditional simulation methods.

2. DATASET DESCRIPTION

The dataset utilized in this study, sourced from Kaggle, includes eleven features of metamaterial antennas and comprises 572 records with frequency range of 2 to 3 GHz [14]. Each record provides comprehensive details about the antennas, such as the gap between rings ($W0m$), height and width of the split ring resonator (SRR) (Wm), distance between rings (Dm), width of rings (Tm), number of SRR cells in the array ($rows$), distance between the antenna patch and the array (Xa), power accepted by the antenna ($p0$), distance between SRR cells in the array (Ya), power radiated by the antenna (pr), gain, bandwidth, and return loss (S_{11}). Table 1 summarizes the characteristics of the dataset, including the minimum, maximum, standard deviation, and mean values for each feature. These statistics provide a comprehensive overview of the data distribution and variability, which are critical for understanding the dataset's structure and for informing the ML models used in the study. The features will be used to predict the antenna's gain, S_{11}, and bandwidth using various regression algorithm.

3. MACHINE LEARNING TECHNIQUES

The rapid advancements in wireless communication systems have led to an increasing demand for compact antennas that maintain wide frequency bands and high gain. Traditional antenna design methods, while being effective, often fall short due to limitations such as low efficiency, low gain, and narrow bandwidth. To address these challenges, recent research has focused on leveraging ML to optimize antenna design and performance. Machine learning offers powerful tools to model and predict complex relationships within datasets, making it an ideal solution for enhancing antenna design. In this context, four popular ML regression models — Extra Trees, Random Forest, XGBoost, and CatBoost — are applied to predict critical antenna parameters, such as the S_{11} parameter, gain, and bandwidth. Extra Trees, or Extremely Randomized Trees, is an ensemble learning method that constructs multiple decision trees during training. Unlike Random Forests, Extra Trees model uses random splits for each node, which can reduce variance but might increase bias. This approach is recognized for its effectiveness and precision in managing large datasets and its capability to model intricate relationships [15]. Random Forest builds multiple decision trees and merges their results to improve predictive performance and control overfitting. By using a bootstrap aggregating (bagging) method and random feature selection for each tree, Random Forests reduce variance and enhance generalization, making them robust against overfitting [16]. XGBoost is a gradient boosting framework known for its speed and performance. It sequentially builds trees, where each tree corrects the errors of the previous ones. XGBoost includes regularization techniques to avoid overfitting and allows for parallel processing, making it extremely efficient for handling large-scale datasets [17]. CatBoost is a gradient boosting algorithm uniquely designed to manage categorical features effectively. It converts categorical data into numerical values internally, which reduces the preprocessing steps required. CatBoost is known for its high accuracy, fast learning, and ability to avoid overfitting, making it suitable for a wide range of regression and classification tasks [18]. To assess the effectiveness of these machine learning models, a range of metrics are employed [19, 20].

Mean Squared Error quantifies the average squared deviation between the actual and predicted values. A lower MSE value denotes the superior model performance, reflecting smaller errors. The calculation is as follows:

$$MSE = \frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{1}$$

Mean Absolute Error (MAE) represents the average absolute difference between actual and predicted values. MAE is straightforward to interpret and provides a measure of prediction accuracy without penalizing large errors more than small ones. It is given by:

$$\text{MAE} = \frac{1}{n}\sum\nolimits_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

R-squared represents the percentage of variance in the dependent variable that can be explained by the independent variables. R^2 values span from 0 to 1, with higher values indicating greater predictive accuracy. It is computed as:

$$R^2 = 1 - \frac{\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum\nolimits_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

Root Mean Squared Error is derived from the square root of the MSE and quantifies the average magnitude of prediction errors. A lower RMSE signifies superior model performance. It is expressed as:

$$RMSE = \sqrt{\frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4}$$

Mean Absolute Percentage Error calculates the average absolute percentage deviation between actual and predicted values. MAPE offers a standardized metric for prediction accuracy, facilitating comparisons across various datasets. It is calculated as:

$$\text{MAPE} = \frac{100\%}{n}\sum\nolimits_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{5}$$

4. RESULTS AND DISCUSSION

Figures 1, 2, and 3 present scatter plots comparing actual versus predicted values of S_{11}, gain, and bandwidth for metamaterial antennas, respectively, using four different machine learning regression models: Extra Trees, Random Forest, XGBoost, and CatBoost. Each model is evaluated under three different configurations with varying sizes of training and testing sets (30%, 50%, and 70% for testing sets). The diagonal dashed line in each subplot represents the ideal prediction line where predicted values perfectly match the actual values. Proximity to this line indicates higher predictive accuracy. The Extra Trees model (Figures 1(a), 2(a), 3(a)) shows a high degree of accuracy across all configurations, with most points clustering closely around the diagonal line. For the 30% testing set, the predictions are nearly ideal, indicating that the model generalizes well with a larger training set. As the testing set size increases to 50% and 70%, the model's predictions remain robust, although slight deviations from the diagonal line become more apparent. This suggests that while the Extra Trees model maintains high accuracy, larger testing set sizes introduce more variability, as expected. The Random Forest model (Figures 1(b), 2(b), 3(b)) also demonstrates strong predictive performance, with points closely aligning with the diagonal line across all testing set sizes. The 30% testing set configuration shows the highest accuracy, with minimal spread around the line. Increasing the testing set size to 50% and 70% results in a modest increase in prediction error, but the overall accuracy remains high. This indicates that the RF model is effective at handling different training/testing splits and maintains robustness in its predictions. XGBoost (Figures 1(c), 2(c), 3(c)) shows excellent predictive accuracy, particularly with the 30% and 50% testing sets, where the points are tightly clustered around the diagonal line. The 70% testing set configuration introduces more spread, indicating a slight reduction in accuracy with a larger testing set. Despite this, XGBoost performs consistently well, highlighting its strength in handling complex data and making accurate predictions even with substantial testing set sizes.

The CatBoost model (Figures 1(d), 2(d), 3(d)) performs well, with predictions closely aligning with the actual values across all configurations. The 30% and 50% testing sets show high accuracy, with points tightly clustered around the diagonal line. The 70% testing set configuration exhibits a greater spread, similar to the other models, indicating increased variability with a larger testing set. However, CatBoost maintains a high level of accuracy, demonstrating its capability to make reliable predictions even with different training/testing splits. Overall, the results across Figures 1, 2, and 3 indicate that as the size of the testing set increases, the variability in predictions also in-

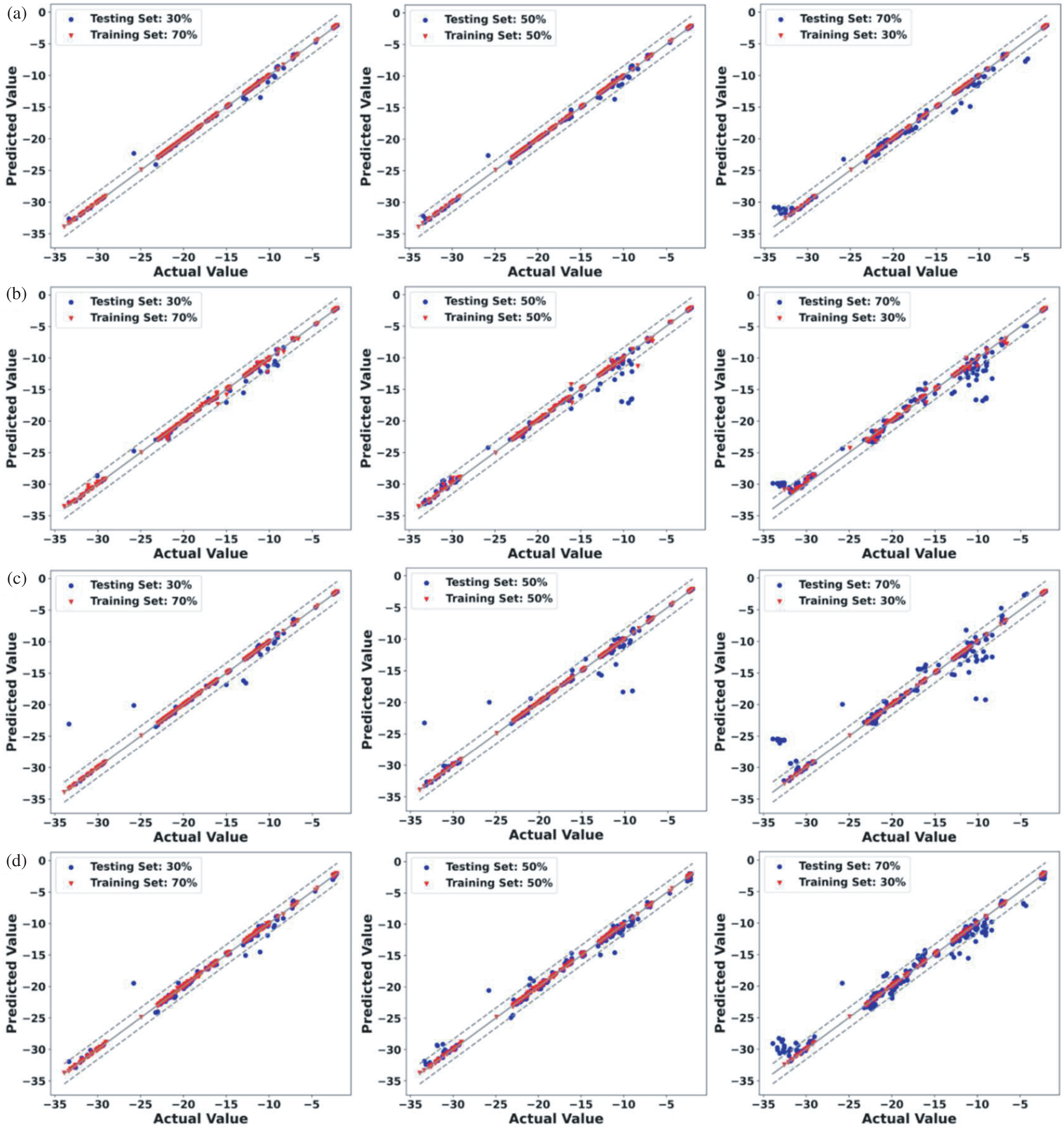

FIGURE 1. Scatter plots of actual versus predicted values of S_{11} for four different machine learning regression models: (a) Extra Trees, (b) Random Forest, (c) XGBoost, and (d) CatBoost.

creases, which is a common trend in machine learning. The Extra Trees and Random Forest models show slightly better performance in terms of predictive accuracy and robustness than XGBoost and CatBoost. This is evident from the tighter clustering of points around the diagonal line, particularly in the configurations with smaller testing sets. These results emphasize the importance of selecting an appropriate training/testing split and choosing a robust model that can handle variability in the data. The consistency of the Extra Trees and Random Forest models suggests that ensemble methods with decision trees are particularly effective for predicting the S_{11}, gain, and bandwidth of metamaterial antennas. The findings from this study offer valuable insights into the advantages and limitations of various machine learning models in this application, guiding future research and model selection for similar predictive tasks.

4.1. Performance Metrics

Table 2 presents a performance comparison of the Extra Trees, XGBoost, CatBoost, and Random Forest regressors on the S_{11} dataset. The Extra Tree Regressor demonstrates superior predictive performance with the highest R^2 values and the lowest error metrics across all test sizes. For instance, at a 30% test size, the Extra Tree Regressor achieves an R^2 of 0.9984, indicating excellent predictive accuracy, and maintains relatively low MSE (0.1032), MAE (0.1128), RMSE (0.3213), and MAPE (0.9082) values compared to the other models. In contrast, the XGBoost regressor shows the lowest performance at larger test sizes, with the highest MSE (2.8081), MAE (0.6288), RMSE (1.6757), and MAPE (4.4208) at a 70% test size. The performance of all models generally decreases as the test size increases. This trend is expected because larger test sets typically introduce more variability and challenge to the model's predictive capabilities. For example, the Extra Tree regressor's R^2 decreases from 0.9984 at a 30% test size to 0.9948 at a 70% test size, while its MSE increases correspondingly. Similar trends are observed for the XGBoost, CatBoost, and Random Forest regressors, with increasing test sizes leading to higher error metrics and lower R^2 values.

TABLE 2. Comparative performance metrics of Extra Trees, XGBoost, CatBoost, and Random Forest regressors applied to the S_{11} dataset, with detailed statistics including MAE, R^2, MSE, RMSE, and MAPE for test sizes of 30%, 50%, and 70%.

	MSE	MAE	R^2	RMSE	MAPE
Test Size	**Extra Tree**				
30%	0.1032	0.1128	0.9984	0.3213	0.9082
50%	0.1772	0.1263	0.9970	0.4210	1.0429
70%	0.3315	0.2210	0.9948	0.5757	1.7457
	XGBoost				
30%	1.0165	0.2462	0.9830	1.0082	1.5905
50%	1.4500	0.3169	0.9786	1.2042	2.5107
70%	2.8081	0.6288	0.9557	1.6757	4.4208
	CatBoost				
30%	0.4592	0.2962	0.9923	0.6777	2.5006
50%	0.3912	0.3452	0.9942	0.6255	3.2259
70%	1.0761	0.5897	0.9830	1.0374	4.7913
	Random Forest				
30%	0.2570	0.2151	0.9957	0.5070	1.7304
50%	1.3396	0.3581	0.9802	1.1574	3.0707
70%	1.6536	0.5993	0.9739	1.2859	4.3554

Table 3 presents a performance comparison of the Extra Trees, XGBoost, CatBoost, and Random Forest regressors on the gain dataset. The Extra Tree regressor again demonstrates superior predictive performance with the highest R^2values and the lowest error metrics across all test sizes. At a 30% test size, the Extra Tree Regressor achieves an R^2 of 0.9990, indicating near-perfect predictive accuracy, and maintains relatively low MSE (0.0002), MAE (0.0069), RMSE (0.0145), and MAPE (0.3106) values compared to the other models. In contrast, the CatBoost regressor shows the lowest performance at larger test sizes, with the highest MAPE of 1.0327 and RMSE of 0.0821 at a 70% test size. The performance of all models

TABLE 3. Evaluation of Extra Trees, XGBoost, CatBoost, and Random Forest regressors on the Gain dataset, presenting metrics such as MAE, R^2, MSE, RMSE, and MAPE for various test set configurations of 30%, 50%, and 70%.

	MSE	MAE	R2	RMSE	MAPE
Test Size	**Extra Tree Regressor**				
30%	0.0002	0.0069	0.9990	0.0145	0.3106
50%	0.0003	0.0094	0.9988	0.0176	0.4378
70%	0.0091	0.0177	0.9571	0.0956	0.7037
	XGBoost				
30%	0.0005	0.0125	0.9975	0.0234	0.5272
50%	0.0005	0.0130	0.9979	0.0233	0.6123
70%	0.0150	0.0245	0.9294	0.1225	0.9633
	CatBoost				
30%	0.0010	0.0148	0.9952	0.0322	0.6840
50%	0.0009	0.0156	0.9966	0.0300	0.7856
70%	0.0067	0.0245	0.9683	0.0821	1.0327
	Random Forest				
30%	0.0019	0.0154	0.9915	0.0431	0.6611
50%	0.0017	0.0148	0.9936	0.0409	0.6671
70%	0.0105	0.0247	0.9509	0.1022	0.9939

TABLE 4. Performance comparison of Extra Trees, XGBoost, CatBoost, and Random Forest regressors on the Bandwidth dataset, detailing MSE, MAE, R^2, RMSE, and MAPE metrics across test sizes of 30%, 50%, and 70%.

	MSE	MAE	R2	RMSE	MAPE
Test Size	**Extra Tree**				
30%	1.8545	0.4291	0.9859	1.3618	0.4409
50%	2.1552	0.4825	0.9799	1.4681	0.4682
70%	2.8633	0.5985	0.9791	1.6921	0.5882
	XGBoost				
30%	2.1430	0.5630	0.9837	1.4639	0.5668
50%	2.8583	0.6254	0.9733	1.6907	0.6009
70%	4.4698	0.7993	0.9674	2.1142	0.8370
	CatBoost				
30%	3.8355	0.7534	0.9720	1.9584	0.8504
50%	4.2509	0.8440	0.9676	2.0618	0.8133
70%	4.9121	0.8577	0.9541	2.2163	0.8741
	Random Forest				
30%	2.9770	0.6108	0.9770	1.7254	0.6178
50%	3.0154	0.6246	0.9731	1.7365	0.6722
70%	3.6824	0.7654	0.9722	1.9190	0.7708

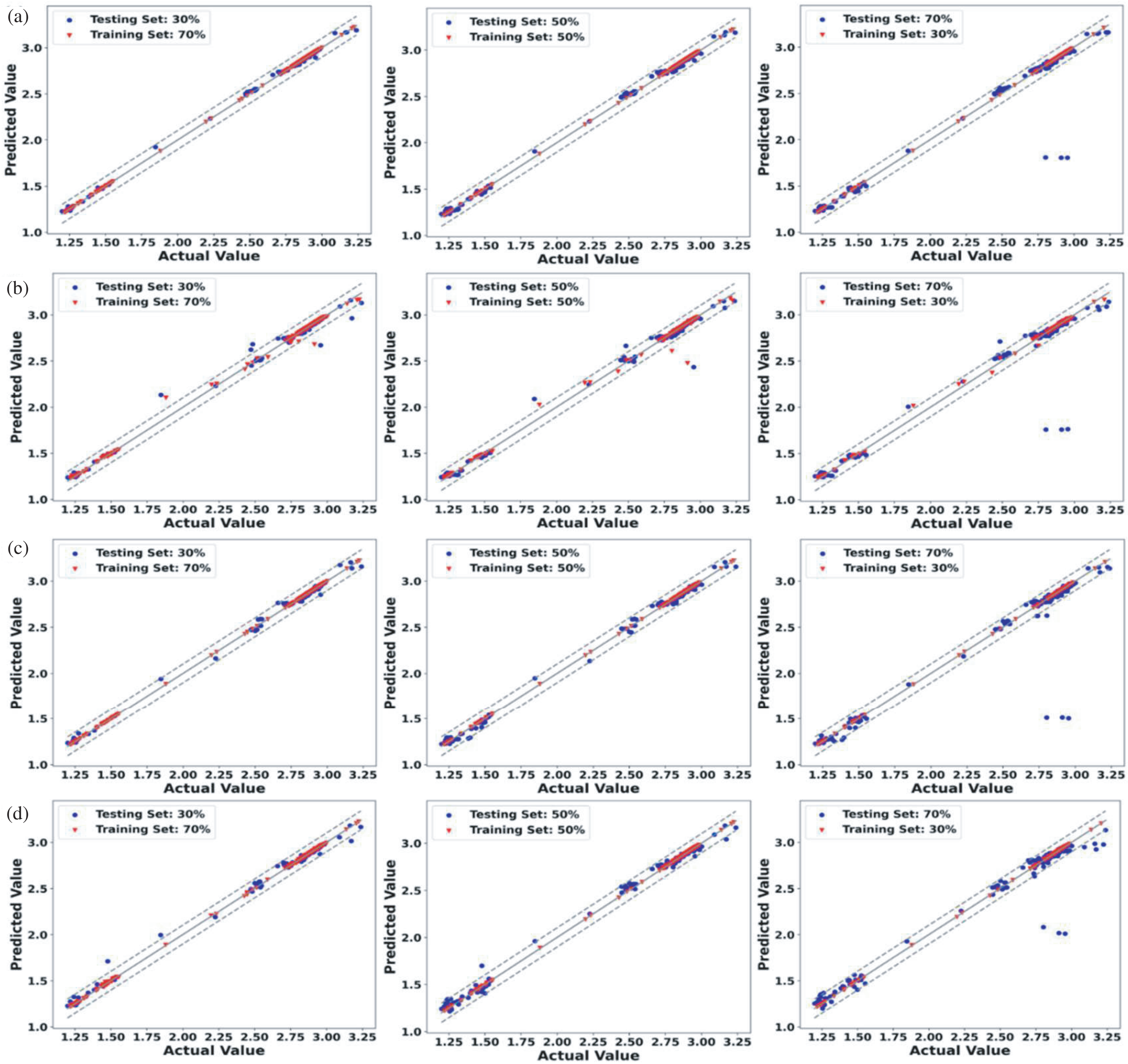

FIGURE 2. Scatter plots demonstrating the relationship between actual versus predicted values of gain, evaluated using four different machine learning regression models: (a) Extra Trees, (b) Random Forest, (c) XGBoost, and (d) CatBoost.

generally decreases as the test size increases. For example, the Extra Tree regressor's R^2 decreases from 0.9990 at a 30% test size to 0.9571 at a 70% test size, while its MSE increases correspondingly. Similar trends are observed for the XGBoost, CatBoost, and Random Forest regressors, with increasing test sizes leading to higher error metrics and lower R^2 values.

Table 4 presents a performance comparison of the Extra Trees, XGBoost, CatBoost, and Random Forest regressors on the bandwidth dataset. The Extra Tree regressor continues to demonstrate superior predictive performance with the highest R^2 values and the lowest error metrics across all test sizes. At a 30% test size, the Extra Tree regressor achieves an R^2 of 0.9859, indicating excellent predictive accuracy, and maintains relatively low MSE (1.8545), MAE (0.4291), RMSE (1.3618), and MAPE (0.4409) values compared to the other models. In contrast, the CatBoost regressor shows the lowest performance, with the highest MSE of 4.9121, MAE of 0.8577, RMSE of 2.2163, and MAPE of 0.8741 at a 70% test size, indicating higher prediction errors and lower accuracy. The performance of all models generally decreases as the test size increases. For example, the Extra Tree regressor's R^2 decreases from 0.9859 at a 30% test size to 0.9791 at a 70% test size, while its MSE

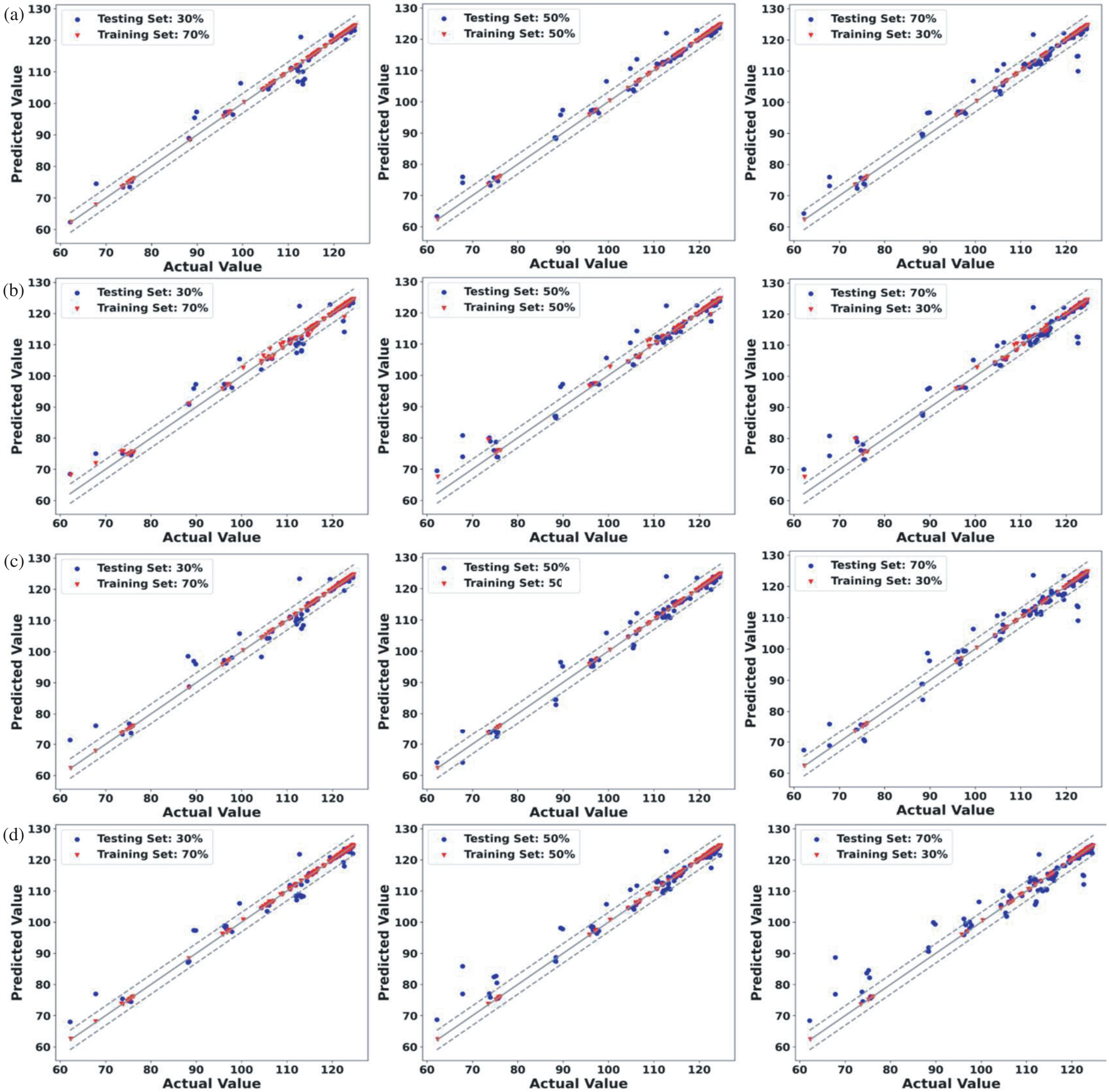

FIGURE 3. Scatter plots demonstrating the relationship between actual versus predicted values of bandwidth, evaluated using four different machine learning regression models: (a) Extra Trees, (b) Random Forest, (c) XGBoost, and (d) CatBoost.

increases correspondingly. Similar trends are observed for the XGBoost, CatBoost, and Random Forest regressors, with increasing test sizes leading to higher error metrics and lower R^2 values.

The Extra Tree regressor consistently outperforms the other models across all datasets due to its ability to reduce variance without significantly increasing bias, a characteristic of ensemble learning methods that construct multiple decision trees during training. This method's robustness against overfitting and its efficient handling of large datasets contribute to its superior performance in predicting the S_{11}, gain, and bandwidth of metamaterial antennas. The results indicate that while ensemble methods like XGBoost and Random Forest also perform well, Extra Trees' specific approach to random splits and aggregation of results yields better accuracy and lower errors, making it the most effective model for these applications. The CatBoost regressor, despite being designed for handling cate-

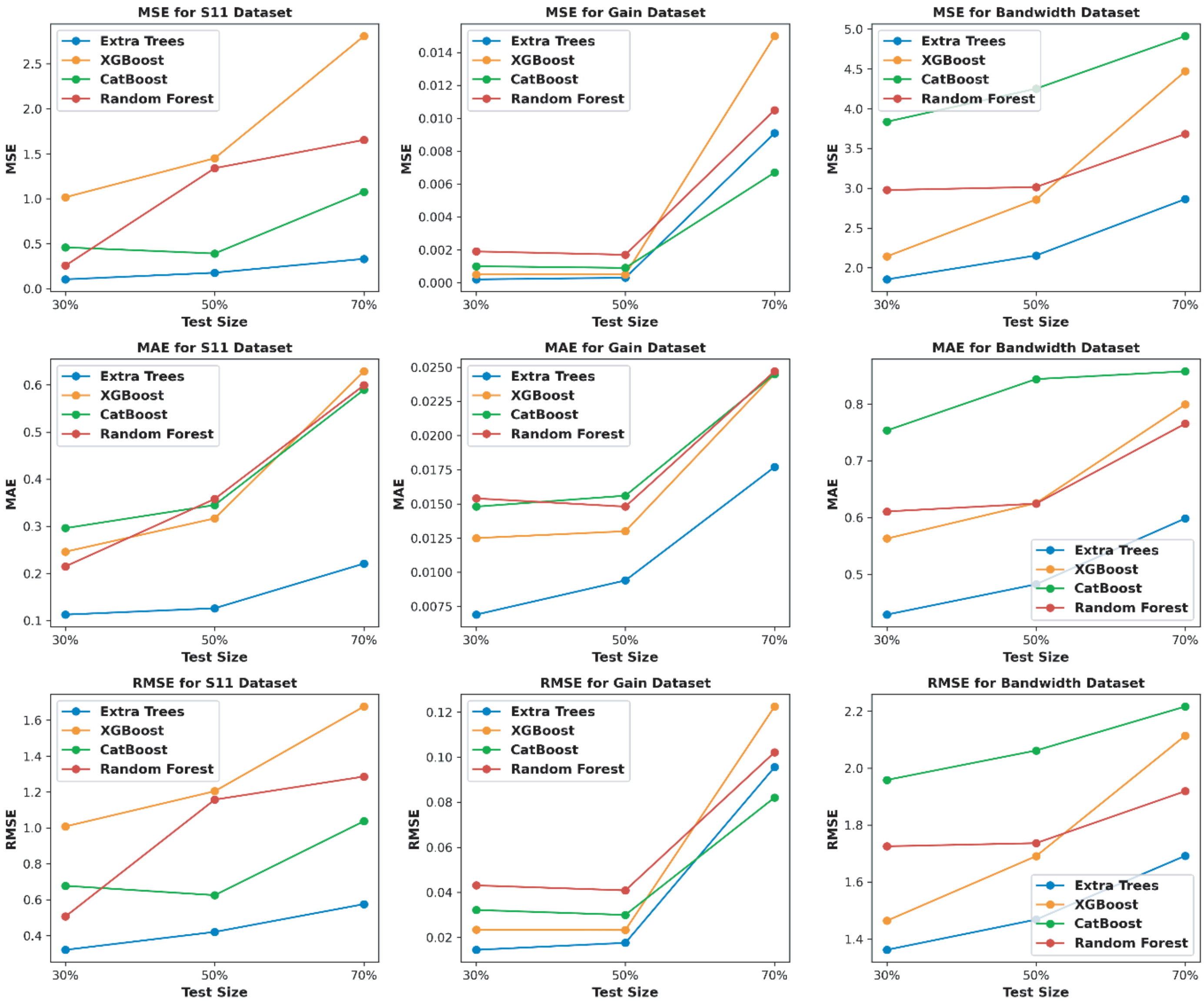

FIGURE 4. Error variation curves for different machine learning models on S_{11}, gain, and bandwidth datasets.

gorical features efficiently, shows lower performance in these contexts, highlighting the importance of model selection based on specific dataset characteristics and prediction tasks.

The processing time for each model was measured using the same computational resources to ensure consistency. Processing time is a crucial factor in evaluating the practicality and efficiency of machine learning models, especially when they are deployed in real-world applications where computational resources and time are limited. The training time refers to the duration required to train the model, while the prediction time is the time taken for the model to make predictions on new data. Shorter times are generally preferable, particularly in scenarios requiring frequent updates or real-time decision-making. From Table 5, we observe that Extra Trees has the shortest training time at 0.1641 seconds and a prediction time of 0.0108 seconds, making it highly efficient. Random Forest also shows efficiency with a training time of 0.2202 seconds and a prediction time of 0.0101 seconds. XGBoost, while being more robust and accurate, takes longer with a training time of 0.5942 seconds and a prediction time of 0.0283 seconds. CatBoost has the longest training time at 0.8215 seconds but offers the fastest prediction time of 0.0069 seconds, suitable for scenarios requiring quick predictions after offline training. These processing times provide insight into the computational efficiency of each model, helping to inform decisions about their deployment in various practical applications. Understanding the trade-offs among training time, prediction time, and model accuracy is essential for optimizing performance in real-world settings.

The results presented in Figure 4 illustrate the performance of four different machine learning models — Extra Trees, XGBoost, CatBoost, and Random Forest—on three antenna parameter datasets: S_{11}, gain, and bandwidth. The analysis focuses on three key error metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error

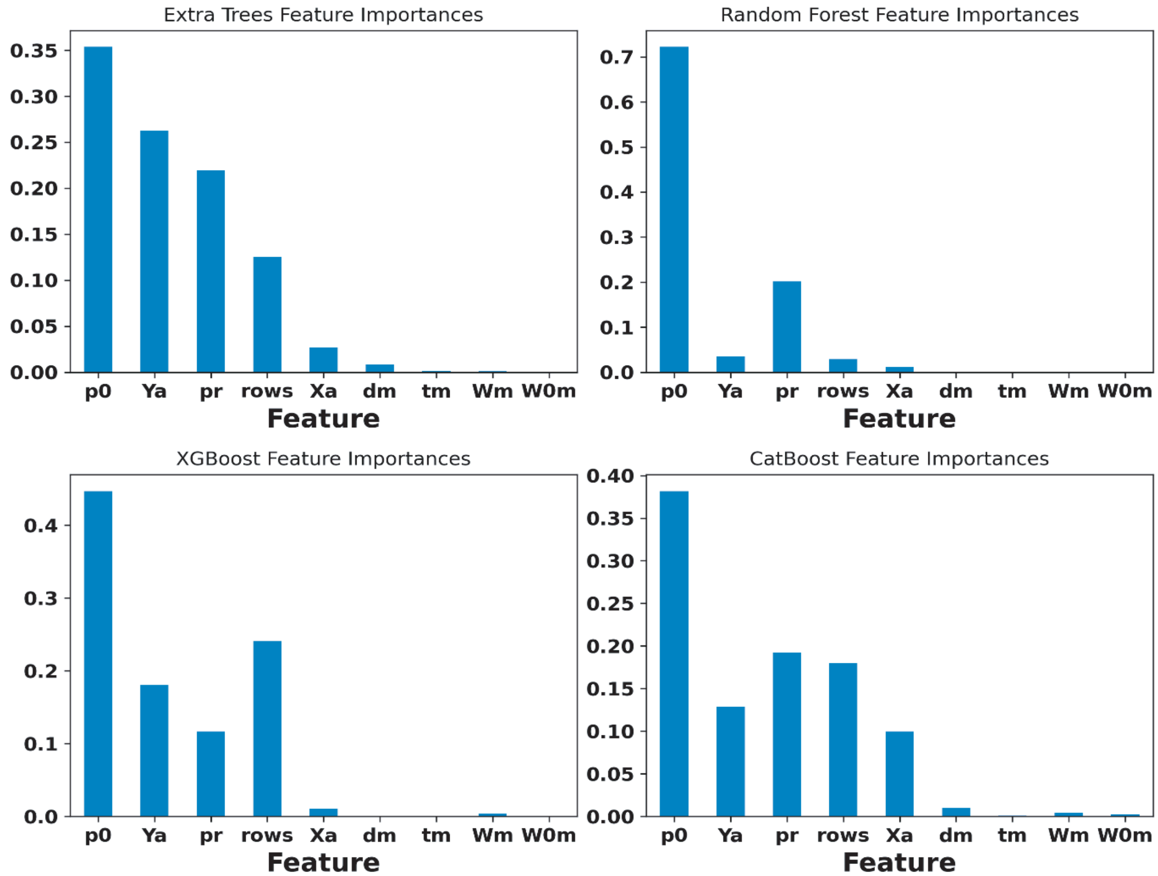

FIGURE 5. Feature importances for metamaterial antenna S_{11} prediction using different machine learning models.

TABLE 5. Processing time for machine learning models.

Model	Training Time (seconds)	Prediction Time (seconds)
Extra Trees	0.1641	0.0108
Random Forest	0.2202	0.0101
XGBoost	0.5942	0.0283
CatBoost	0.8215	0.0069

(RMSE), evaluated across varying test sizes (30%, 50%, 70%). For the S_{11} dataset, Extra Trees consistently demonstrates the lowest MSE across all test sizes, indicating high accuracy. XGBoost and Random Forest, on the other hand, exhibit a sharp increase in error as the test size increases, suggesting potential overfitting issues with smaller training data. Extra Trees also maintains the lowest MAE values, closely followed by CatBoost, while XGBoost and Random Forest display higher MAE values, particularly at larger test sizes. The RMSE metric shows a similar trend to MSE, with Extra Trees achieving the lowest values, reflecting their robustness and predictive accuracy for the S_{11} dataset. In contrast, XGBoost and Random Forest perform worse at higher test sizes. In the case of the gain dataset, all models show relatively low MSE values, with Extra Trees and XGBoost performing the best. CatBoost and Random Forest demonstrate slightly higher MSE but remain competitive. The MAE values indicate that Extra Trees and XGBoost outperform the other models, maintaining lower error rates across all test sizes, while CatBoost and Random Forest have higher MAE, particularly at larger test sizes. The RMSE values reinforce these trends, with Extra Trees and XGBoost maintaining lower values, confirming their superior performance on the gain dataset, while CatBoost and Random Forest exhibit increased RMSE at higher test sizes. For the bandwidth dataset, Extra Trees exhibit the lowest MSE, followed by Random Forest. XGBoost and CatBoost show higher MSE, with CatBoost performing the worst as the test size increases. Extra Trees also consistently achieves the lowest MAE, indicating high predictive accuracy, with Random Forest following and XGBoost and CatBoost having higher MAE values. The RMSE values align with the observed MSE and MAE trends, with Extra Trees leading in performance, Random Forest remaining competitive, and XGBoost and CatBoost lagging behind.

4.2. Feature Importance

Feature importance is a key metric in understanding the impact of each feature on the prediction models. In this study, we compute feature importance for each machine learning model using the following methods:

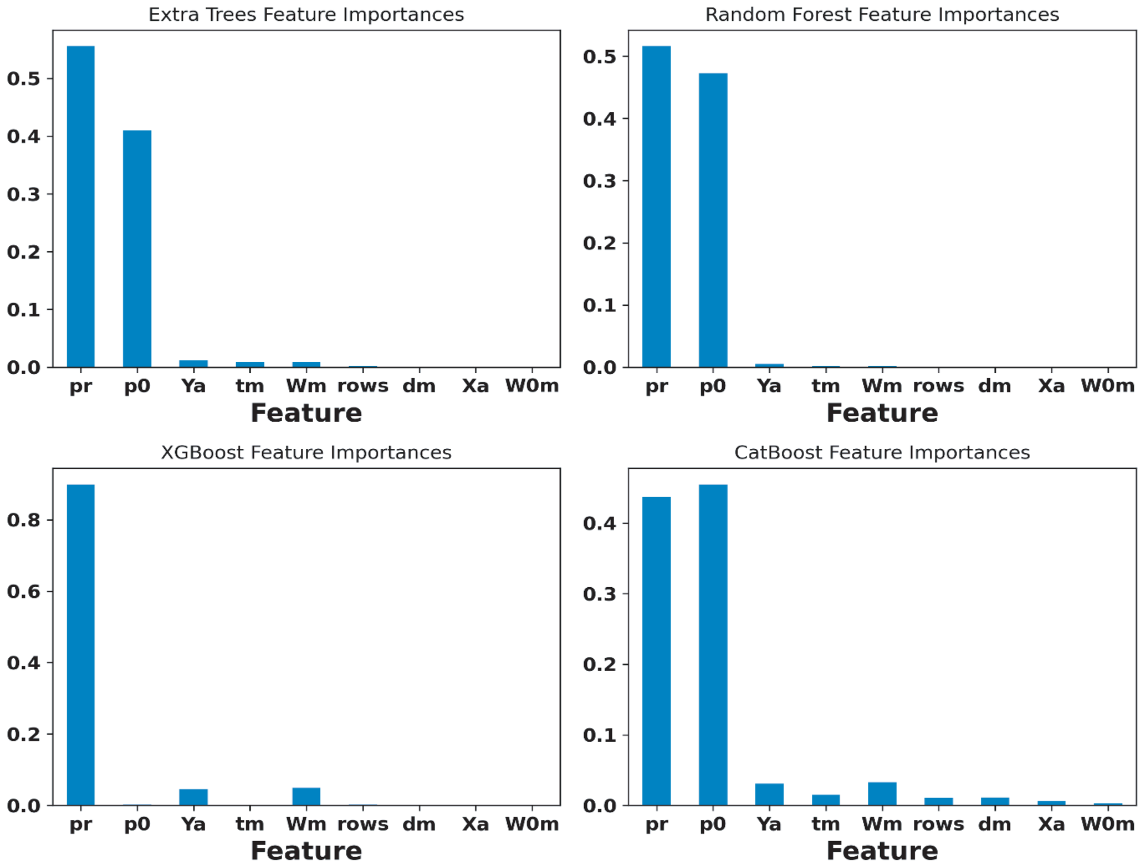

FIGURE 6. Feature importances for metamaterial antenna gain prediction using different machine learning models.

Feature importance in tree-based models like Extra Trees and Random Forest is calculated based on the average decrease in impurity (Gini impurity or entropy) across all trees. The importance of a feature f is computed as:

$$FI\left(f\right)=\sum_{t=1}^{T}\left(\frac{1}{N_t}\sum_{n=1}^{N_t}\Delta I_n(f)\right) \tag{6}$$

where T is the total number of trees, N_t the number of nodes in tree t, and $\Delta I_n(f)$ the decrease in impurity caused by feature f at node n.

XGBoost computes feature importance using the gain, cover, and frequency methods. The gain method, which measures the improvement in accuracy brought by a feature to the branches it is on, is calculated as:

$$FI\left(f\right)=\sum_{i\in nodes(f)}\frac{G_i}{H_i} \tag{7}$$

where G_i and H_i are the sum of gradient and hessian statistics for the splits where feature f is used.

CatBoost uses a permutation-based method to calculate feature importance. It measures the change in the loss function after permuting the feature's values. The importance of a feature f is given by:

$$FI\left(f\right)=E_\pi[L\left(y,\,\hat{y}_{\pi(f)}\right)-L(y\hat{y}) \tag{8}$$

where L is the loss function, $\hat{y}$ the predicted values, and $\pi(f)$ the permutation of feature f. These calculations help in identifying the most influential features, which is crucial for model interpretation and improvement.

Figures 5, 6, and 7 illustrate the feature importances for predicting the S_{11}, gain, and bandwidth of metamaterial antennas, respectively, using four different machine learning regression models: Extra Trees, Random Forest, XGBoost, and CatBoost. Each subplot represents one model, showing the relative importance of each feature in the dataset. Figure 5 shows that the feature $p0$ consistently has the highest importance across all models, with values of 0.3501 for Extra Trees, 0.7176 for Random Forest, 0.4468 for XGBoost, and 0.3818 for CatBoost. This indicates that $p0$ is the most significant predictor of the S_{11} parameter. The feature Ya is the second most important for Extra Trees (0.2848) and XGBoost (0.1807), but it has lower importance for Random Forest (0.0359) and CatBoost (0.1290). Similarly, pr has moderate importance across all models, with values of 0.1831 for Extra Trees, 0.2046 for Random Forest, 0.1169 for XGBoost, and 0.1920 for CatBoost. The number of split ring resonator cells ($rows$) shows significant importance for XGBoost (0.2410) and moderate importance for CatBoost (0.1799), but it is less important for Extra Trees (0.1327) and Random Forest (0.0293). The feature Xa, while generally having lower importance, is relatively more significant for CatBoost (0.0998) than the other models. Features dm, tm,

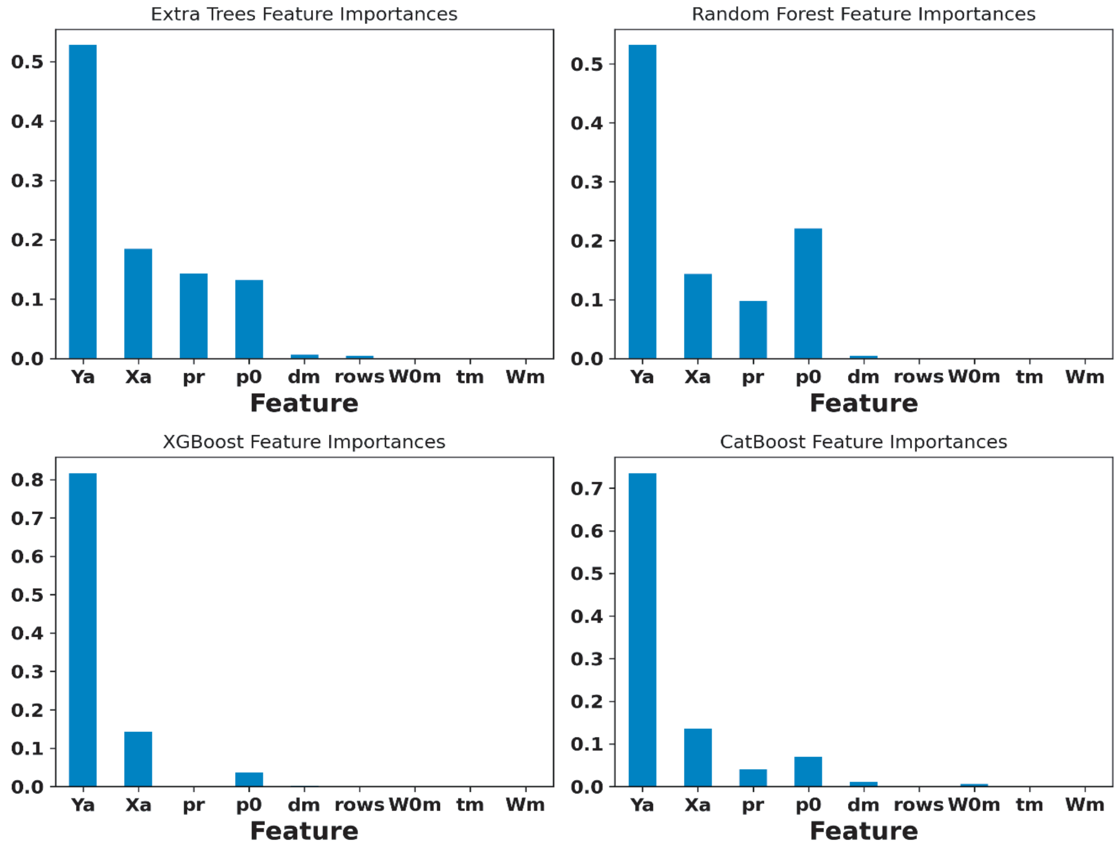

FIGURE 7. Feature importances for metamaterial antenna bandwidth prediction using different machine learning models.

Wm, and $W0m$ have minimal importance across all models, indicating that they have little influence on predicting the S_{11} parameter. The feature importances are normalized to a scale from 0 to 1 for consistency across models. This visualization helps identify which features have the most significant impact on predicting the S_{11} parameter, guiding further refinement and optimization of the model.

Figure 6 shows that the feature pr has the highest importance for Extra Trees (0.5167), Random Forest (0.5617), and XGBoost (0.8991), and it is also significantly important for CatBoost (0.4372). This indicates that pr is a critical predictor of the antenna gain across all models. The feature $p0$ is highly important for Extra Trees (0.4511), Random Forest (0.4270), and CatBoost (0.4548), but it shows minimal importance for XGBoost (0.0025). The feature Ya has lower importance overall, with values of 0.0117 for Extra Trees, 0.0057 for Random Forest, 0.0449 for XGBoost, and 0.0304 for CatBoost. The feature Wm is relatively more important for XGBoost (0.0490) and CatBoost (0.0327), but it is less important for Extra Trees (0.0081) and Random Forest (0.0016). The feature tm has low importance across all models, with values of 0.0077 for Extra Trees, 0.0022 for Random Forest, 0.0000 for XGBoost, and 0.0147 for CatBoost. The feature rows show minimal importance across all models, with values of 0.0018 for Extra Trees, 0.0005 for Random Forest, 0.0020 for XGBoost, and 0.0104 for CatBoost. Similarly, the feature dm has low importance, with values of 0.0014 for Extra Trees, 0.0009 for Random Forest, 0.0012 for XGBoost, and 0.0113 for CatBoost. The feature Xa has minimal importance, with values of 0.0012 for Extra Trees, 0.0004 for Random Forest, 0.0010 for XGBoost, and 0.0058 for CatBoost. The feature $W0m$ has the lowest importance across all models, with values of 0.0002 for Extra Trees, 0.0001 for Random Forest, 0.0003 for XGBoost, and 0.0027 for CatBoost.

Figure 7 shows that the feature Ya has the highest importance for all models: 0.5239 for Extra Trees, 0.5337 for Random Forest, 0.8170 for XGBoost, and 0.7356 for CatBoost. This indicates that Ya is a critical predictor of the antenna bandwidth across all models. The feature $p0$ is also important but to a lesser degree, with values of 0.1588 for Extra Trees, 0.2018 for Random Forest, 0.0360 for XGBoost, and 0.0703 for CatBoost. The feature Xa has significant importance as well, with values of 0.1559 for Extra Trees, 0.1594 for Random Forest, 0.1428 for XGBoost, and 0.1355 for CatBoost. The feature pr has moderate importance for Extra Trees (0.1489) and Random Forest (0.0986), but it shows minimal importance for XGBoost (0.0008) and CatBoost (0.0399). Other features such as dm, $rows$, $W0m$, tm, and Wm have very low importance across all models, indicating that they have little influence on predicting the bandwidth of the antenna. The feature importances are normalized to a scale from 0 to 1 for consistency across models. This visualization helps identify which features have the most significant impact on predicting the antenna bandwidth, guiding further refinement and optimization of the model.

5. CONCLUSIONS

This study validates the effectiveness of ML regression models in predicting key antenna parameters, offering a viable alternative to traditional design methods. By evaluating Extra Trees, Random Forest, XGBoost, and CatBoost across different training and testing set sizes, models that consistently deliver high predictive accuracy for S_{11}, gain, and bandwidth have been identified. Specifically, the Extra Trees model outperforms other models, achieving R^2values of 0.9984 for S_{11}, 0.9990 for gain, and 0.9859 for bandwidth. Feature importance analysis revealed that specific features, such as pr and $p0$, play a critical role in the predictive models, providing valuable insights into the parameters that most significantly impact antenna performance. Machine learning approaches are increasingly significant in contemporary research and are anticipated to play a critical role in future technological advancements. The accuracy of predictions largely depends on the chosen model, as evidenced by the findings. Ensemble methods like Extra Trees and XGBoost, which integrate multiple weak learners, have shown superior performance, especially when being combined with large datasets. This research underscores the potential of machine learning to streamline antenna development processes, making them more efficient and cost-effective. Future work may explore additional machine learning algorithms and feature engineering techniques to further enhance predictive performance. Integrating these models into a comprehensive design framework could significantly expedite the antenna development cycle. Furthermore, comparing the performance of these machine learning models with traditional simulation tools like CST software in future studies could provide deeper insights into their practical applicability and robustness.

ACKNOWLEDGEMENT

The authors wish to express their profound gratitude to Universiti Teknikal Malaysia Melaka (UTeM) for their generous support. This work was made possible through the grant PJP/2024/FTKEK/PERINTIS/S01388. Their assistance and resources have been instrumental in the successful completion of this research.

REFERENCES

[1] Hong, T., C. Liu, and M. Kadoch, "Machine learning based antenna design for physical layer security in ambient backscatter communications," *Wireless Communications and Mobile Computing*, Vol. 2019, No. 1, 4870656, 2019.

[2] Zhang, J., J. Xu, Q. Chen, and H. Li, "Machine-learning-assisted antenna optimization with data augmentation," *IEEE Antennas Wirel. Propag. Lett.*, Vol. 22, No. 8, 1932–1936, 2023.

[3] Li, W. T., H. S. Tang, C. Cui, Y. Q. Hei, and X. W. Shi, "Efficient online data-driven enhanced-XGBoost method for antenna optimization," *IEEE Trans. Antennas Propag.*, Vol. 70, No. 7, 4953–4964, 2022.

[4] Verma, R. K. and D. K. Srivastava, "Bandwidth enhancement of a slot loaded T-shape patch antenna," *Journal of Computational Electronics*, Vol. 18, 205–210, 2019.

[5] Kumar, A. and M. K. Singh, "Band-notched planar UWB microstrip antenna with T-shaped slot," *Radioelectronics and Communications Systems*, Vol. 61, No. 8, 371–376, 2018.

[6] Geetharamani, G. and T. Aathmanesan, "Design of metamaterial antenna for 2.4 GHz WiFi applications," *Wireless Personal Communications*, Vol. 113, 2289–2300, 2020.

[7] Yan, S., P. J. Soh, and G. A. E. Vandenbosch, "Compact all-textile dual-band antenna loaded with metamaterial-inspired structure," *IEEE Antennas and Wireless Propagation Letters*, Vol. 14, 1486–1489, 2014.

[8] Rawal, K., P. D. Devendrabhai, P. Pataniya, P. Jain, A. Joshi, G. K. Solanki, and M. Tannarana, "Versatile photo-sensing ability of paper based flexible 2D-Sb0. 3Sn0. 7Se2 photodetector and performance prediction with machine learning algorithm," *Optical Materials*, Vol. 152, 115547, 2024.

[9] Watpade, A. D., S. Thakor, P. Jain, P. P. Mohapatra, C. R. Vaja, A. Joshi, D. V. Shah, and M. T. Islam, "Comparative analysis of machine learning models for predicting dielectric properties in MoS2 nanofiller-reinforced epoxy composites," *Ain Shams Engineering Journal*, Vol. 15, No. 6, 102754, 2024.

[10] Jain, P., A. Joshi, and A. Joshi, "Assessing the efficacy of machine learning models in hydroxyapatite nano-powder assisted electro discharge machining of Ti-6Al-4 V Grade-5 alloy," *International Journal on Interactive Design and Manufacturing (IJIDeM)*, 1–9, 2024.

[11] Song, K., F. Yan, T. Ding, L. Gao, and S. Lu, "A steel property optimization model based on the XGBoost algorithm and improved PSO," *Computational Materials Science*, Vol. 174, 109472, 2020.

[12] Watpade, A. D., S. Thakor, P. Sharma, D. V. Shah, C. R. Vaja, and P. Jain, "Synthesis, characterization, and dielectric spectroscopy of TiO_2 and ZnO nanoparticle-reinforced epoxy composites," *Journal of Materials Science: Materials in Electronics*, Vol. 35, No. 7, 466, 2024.

[13] Paulson, G., K. Upadhyay, P. Dighe, S. Pathan, and Tanweer, "Regression analysis of metamaterial antenna using decision and extra tree regressors," in *Proceedings of the 2023 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom)*, 313–316, Dubai, United Arab Emirates, 2023.

[14] Machado, R., "Metamaterial antennas," Available: https://www.kaggle.com/renanmav/metamaterial-antennas, 2019.

[15] Kurniawati, N., D. N. N. Putri, and Y. K. Ningsih, "Random forest regression for predicting metamaterial antenna parameters," in *Proc. of the 2020 2nd Intern. Conf. on Industrial Electrical and Electronics (ICIEE)*, 174–178, Lombok, Indonesia, 2020.

[16] Shingala, B., P. Panchal, S. Thakor, P. Jain, A. Joshi, C. R. Vaja, R. K. Siddharth, and V. A. Rana, "Random forest regression analysis for estimating dielectric properties in epoxy composites doped with hybrid nano fillers," *Journal of Macromolecular Science, Part B*, 1–15, 2024.

[17] Jain, P., J. Yedukondalu, H. Chhabra, U. Chauhan, and L. D. Sharma, "EEG-based detection of cognitive load using VMD and LightGBM classifier," *International Journal of Machine Learning and Cybernetics*, 1–18, 2024.

[18] Dhananjay, B. and J. Sivaraman, "Analysis and classification of heart rate using CatBoost feature ranking model," *Biomedical Signal Processing and Control*, Vol. 68, 102610, 2021.

[19] Abdelhamid, A. A. and S. R. Alotaibi, "Optimized two-level ensemble model for predicting the parameters of metamaterial antenna," *Computers, Materials & Continua*, Vol. 73, No. 1, 917–933, 2022.

[20] El-kenawy, E. M., A. Ibrahim, S. Mirjalili, Y. Zhang, S. Elnazer, and R. M. Zaki, "Optimized ensemble algorithm for predicting metamaterial antenna parameters," *Computers, Materials & Continua*, Vol. 71, No. 3, 4989–5003, 2022.