IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 September 2023, accepted 20 September 2023, date of publication 28 September 2023, date of current version 6 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3320581

RESEARCH ARTICLE

Classification and Prediction of Communication Cables Length Based on S-Parameters Using a Machine-Learning Method

MOHAMMAD AL BATAINEH[1,2], MALIK MOHAMED UMAR[1], AQUIB MOIN[3], MOUSA I. HUSSEIN[1], (Senior Member, IEEE), AND MAHMOUD AL AHMAD[1], (Senior Member, IEEE)

[1]Electrical and Communication Engineering Department, United Arab Emirates University, Al Ain, United Arab Emirates
[2]Telecommunications Engineering Department, Yarmouk University, Irbid 21163, Jordan
[3]Physics Department, United Arab Emirates University, Al Ain, United Arab Emirates

Corresponding author: Mahmoud Al Ahmad (m.alahmad@uaeu.ac.ae)

ABSTRACT The length of the communication cables is a significant indicator of signal integrity. The associated scattering parameters characteristics of a communication cable effectively enable its length estimation. This paper proposes a novel machine learning-based algorithm that utilizes Support Vector Machines (SVM) to predict and classify cable lengths. Specifically, the algorithm employs an SVM Regression Model (SVR) and an SVM Classification Model (SVC) to predict and classify cable lengths based on their S-parameters (S_{21} measurements). As the data under investigation are inseparable, linear, and high-dimensional, SVM has been implemented. The current approach was implemented to verify the length of two datasets, underground and overground cables, with different environmental conditions. The present research introduces an innovative machine-learning algorithm that employs an S-parameter-centric methodology to predict variations in communication cable lengths. Specifically, the SVR model achieved R^2 values of approximately 0.987 for underground cables and 0.991 for overground cables. Meanwhile, the SVC model demonstrated varying levels of accuracy, with optimal performance seen in five classes for underground cables and four classes for overground cables. The SVM model efficiently extracts and weighs features for high-accuracy predictions in nonlinear, multiclass scenarios, making it the optimal model for this work.

INDEX TERMS Classification, prediction, cable length, S-parameters, machine-learning method.

I. INTRODUCTION

The cable access network encompasses many potential problem areas that could be ameliorated through machine learning techniques [1]. Prior to their utilization, the cable's reference number must be ascertained, which can be derived from information imprinted on the outer jacket or available in accompanying data sheets [2]. Nevertheless, these labels may become illegible or unattainable over time or due to other influencing factors, particularly in the context of subterranean cables. Owing to the similarities in appearance among numerous cables, their differentiation through visual inspection proves challenging [3]. Consequently, it is essential to employ prediction/classification methodologies to examine and analyze the cables under consideration [4].

The motivation for this work is to address the cable prediction challenge associated with the Ground-based Radio Array Facility (GRAF) at UAE University, as depicted in Figure 1. The GRAF comprises a 256-element interferometer array designed to operate within the 50–300 MHz frequency range. The antennas, arranged in a pseudorandom configuration, are grouped into 16 clusters, each feeding an analog beamformer. This setup allows the array to track celestial objects through 16 in-field beamformers.

Cable prediction is critical for this setup because determining the length of the underground wire during signal research

The associate editor coordinating the review of this manuscript and approving it for publication was Barbara Masini.

TABLE 1. Dataset format.

No.	Frequency (MHz)	Group X			Group Y		
		Cable Length (m)	S_{21} Magnitude (dB)	S_{21} Phase (degrees)	Cable Length (m)	S_{21} Magnitude (dB)	S_{21} Phase (degrees)
1	0.1	91.27396	-0.7105630	-16.1838980	99.79064	-0.4818216	-0.4818216
2	0.2	91.2597	-0.9925018	-30.5349093	99.49252	-0.7982240	-0.7982240
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
2000	200	90.5991	-8.3162054	22.7280398	100.26060	-9.5099386	-9.5099386

FIGURE 1. The Ground-Based Radio Array Facility (GRAF) at UAE university.

is not always feasible, given that the phased array antenna is 150 meters away from the control room. Therefore, the main challenge will be the need to estimate the length of the installed cables.

Scattering parameters have been used to estimate the communication cable features in several works. The proposed methodologies consider the magnitude and phase of scattering parameters as cable features, in addition to length and frequency. Hollaus et al. have proposed a frequency-based transmission line parameter combined with time domain modeling to extract the effective cable parameters of unshielded conductors [5]. The method utilized transmission (S_{12}) and reflection (S_{11}) scattering parameters to generate the corresponding lumped element models. Bader et al. have demonstrated the use of support vector machine algorithms for the identification of communication cables [6]. They have utilized the S_{11} parameter as predictor features while considering cable length, type, and connectors as classes [7]. They specifically classified 101 frequencies and ten cables using the KNN classifier. The obtained results demonstrated a 99% prediction accuracy of the classifier. Said et al. have proposed a deep learning-based fault classification and location for underground cables. They utilized the one-dimensional convolutional neural network (1D-CNN) and a binary support vector machine (BSVM) to classify the fault type and locate it in real-time scenarios with a 99.6% accuracy [8], [9]. Mishra et al. have presented a tree-based machine-learning technique combined with random forest algorithms to categorize various permanent faults in an underground cable. Their suggested approach provides a rapid and accurate fault classification evaluation of the fault current. 98.8% [10].

This work predicts and classifies the length of communication cables using a machine-learning method. Based on the corresponding S-parameters measurements.

II. S-PARAMETERS MEASUREMENTS

The S-parameter matrix describes the characteristics of the electrical cable as a two-port network [11]. It consists of four parameters: S_{11}, S_{12}, S_{21} and S_{22}. Each parameter has magnitude, and phase provides a relationship between the incident, reflected, and transmitted waves over a specific range of frequencies. These parameters can be used to extract transmission cable characteristics such as permittivity, cable length, and quality. The S-parameters are deemed standard and popular parameters for frequency range analysis, self-reflection, and power transmission between endpoints [11]. Both S_{11} and S_{22} are employed for self-reflection, but their significance may not be sufficient for feature selection. S_{21} or S_{12} are of particular importance as it represents the power transmission from port 1 to port 2. The corresponding scattering parameters for the two sets of cables have been measured using the Rohde and Schwarz ZNL6 - Vector Network Analyzer [12].

III. DATASET PREPARATION

The procedure of dataset preparation, coupled with pertinent feature selection, is crucial in any classification algorithm, as it considerably impacts prediction accuracy. This section comprehensively outlines the approach undertaken for dataset preparation and feature selection.

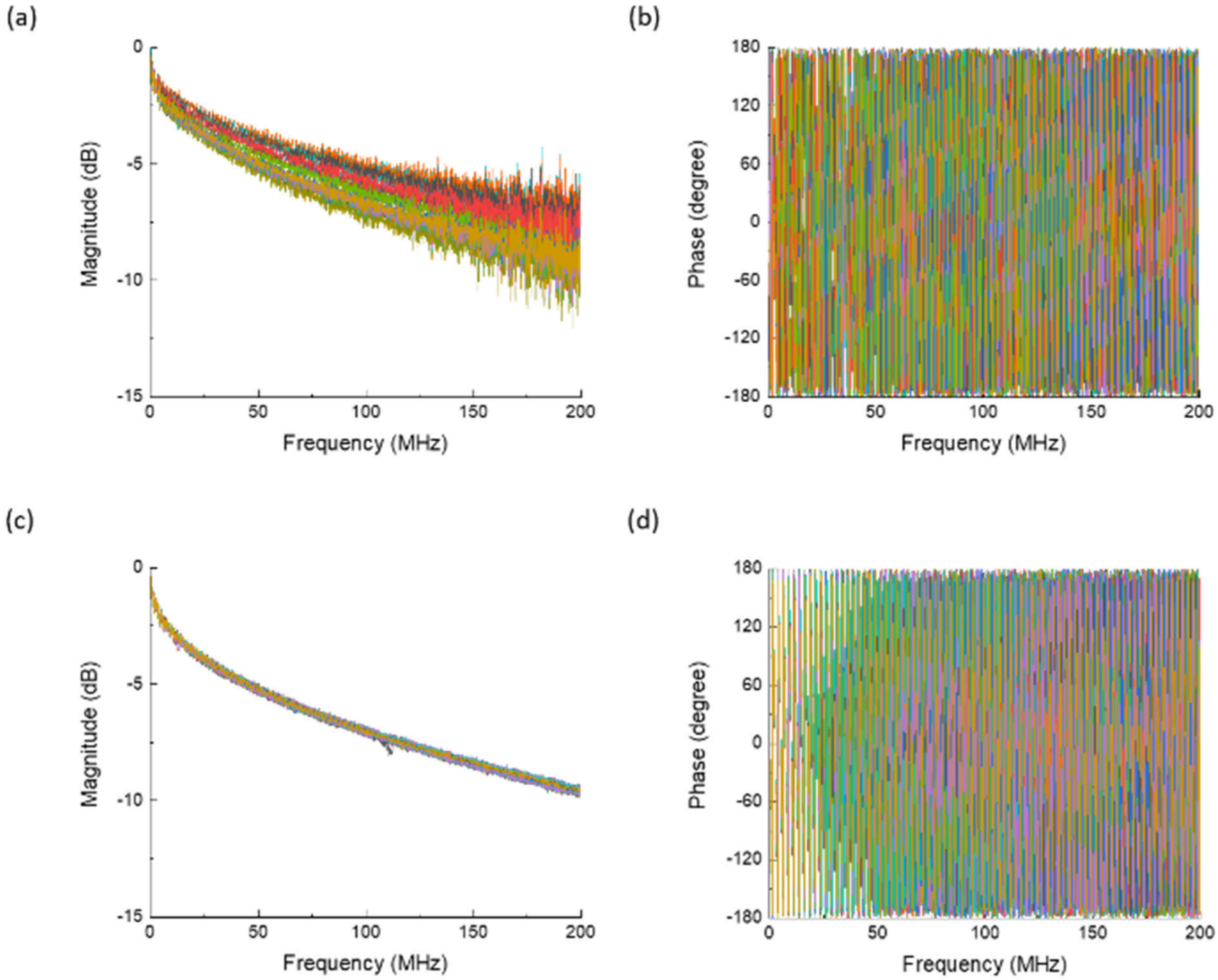

FIGURE 2. Measured S21 parameter. (a) magnitude in dB for cable above ground (Group X). (b) phase in degree for cable over ground (Group X). (c) magnitude in dB for cable underground (Group Y). (d) phase in degree for cable underground (Group Y).

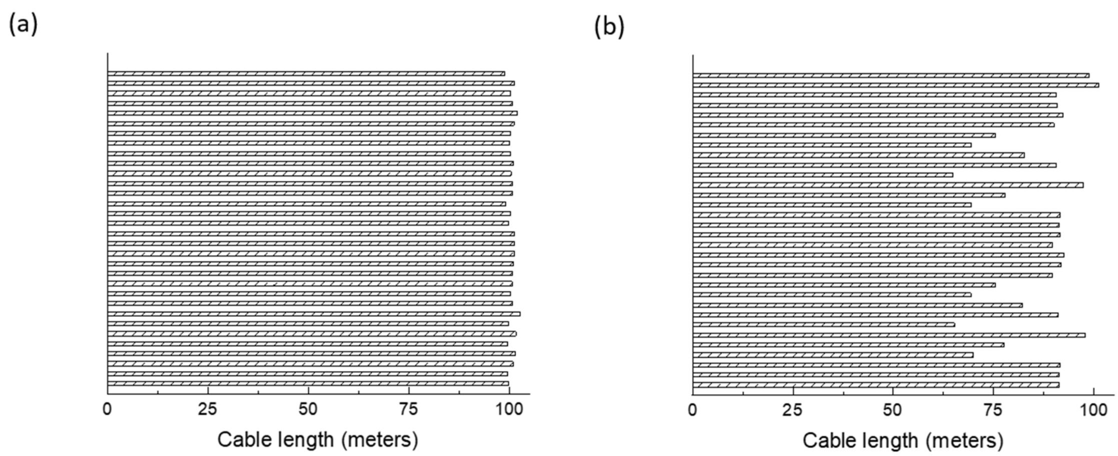

FIGURE 3. Actual cable distribution for (a) Group Y and (b) Group X.

Initially, the frequency range was adjusted from 100 kHz to 200 MHz. Within this range, 2000 frequencies were captured as frequency observations. Subsequently, seven parameters were gathered for the 30 pairs of Group X and Group Y cables under investigation. These seven parameters include the frequency, cable length, magnitude, and phase. Group X comprises 30 underground cables, while Group Y is placed above the ground. Consequently, 60,000 observations pertaining to the seven features based on the four distinct classes specified in Table 1 were collected to construct the necessary dataset. The measured S_{21} parameter versus frequency for both Group X and Group Y is plotted in Figure 2.

Figure 2(a) depicts the magnitude versus frequency for different lengths. The magnitude decreases as the frequency goes up. The phase changes from −180 to 180, as shown in Figure 2(b). Figure 2(c) represents the magnitude versus frequency for different lengths for Group Y. By comparing Figure 2(a) with 2(c), cables in Group Y exhibit higher quality than in Group X because they are protected from harsh environmental conditions. Also, the corresponding measured phase for Group Y, in Figure 2(d), exhibits a more uniform pattern distribution than the one for Group X. The actual length distribution for Group X and Group Y cables are depicted in Figure 3. Group Y cables range between 98.97 to 102.65 meters, while Group X cables range between 64.88 to 97.73 meters.

Table 1 demonstrates that 2,000 distinct frequencies within the range of 100 kHz to 200 MHz were considered for each of the 30 unique pairs of Group X and Group Y cables in this study. As a result, a total of 60,000 observations were gathered. The primary objective of this research is to predict and classify the various lengths of communication cables based on their S-parameter values. Table 2 provides a comprehensive overview of the methodological framework for the machine learning models to be employed in this study. It outlines the two primary models—Support Vector Regression (SVR) and Support Vector Classification (SVC)—along with the specific input features that will be utilized for analysis. Additionally, the table indicates the anticipated output variables for each model and delineates the analytical paradigm under which each operates. This structured representation aims to offer a clear and concise roadmap for the ensuing modeling and simulation activities.

TABLE 2. Features, range of cable lengths, and classes.

Model	Input Features Utilized	Anticipated Output Variable	Analytical Paradigm
Support Vector Regression (SVR)	Scattering Parameter S_{21} (Magnitude & Phase)	Predicted Continuous Cable Length (meters)	Regression Analysis
Support Vector Classification (SVC)	Scattering Parameter S_{21} (Magnitude & Phase)	Predicted Categorical Cable Length (Class Label)	Classification Analysis

IV. PROPOSED SVM MODELS

Support Vector Machines (SVM) is a class of supervised learning algorithms that have gained prominence in classification and regression tasks [13]. The core principle behind SVM is the construction of hyperplanes in a multi-dimensional space that distinctly categorizes the input data points into classes or estimates continuous values in the context of regression [14]. In this work, we propose a novel application of Support Vector Machines (SVM) with a polynomial kernel to predict and classify the physical cable lengths of two distinct groups of communication cables. The prediction and classification tasks are based on the S_{21} measurements (magnitude, phase, and frequency) as outlined in Table 1. Consequently, an SVM Regression Model (SVR) and an SVM Classification model (SVC) are proposed. This section provides a detailed description of the two models.

A. PROBLEM FORMULATION

Given a dataset $(\boldsymbol{x_1}, y_1), (\boldsymbol{x_2}, y_2), \ldots, (\boldsymbol{x_N}, y_N)$, where each $\boldsymbol{x_i}$ denotes an input feature vector (comprising frequency, magnitude, and phase) and y_i is its associated output (cable length). The objective is twofold:

- **Regression**: Predict a continuous value for the cable length.
- **Classification**: Categorize the cables into predefined classes based on specific criteria.

B. SVM REGRESSION (SVR) FORMULATION

Given our dataset, where each $\boldsymbol{x_i}$represents a feature vector encapsulating the S_{21} parameters (magnitude, phase, and frequency) for a specific cable and y_i is its corresponding cable length, the SVM regression model seeks to predict this length by solving the following optimization problem:

$$\min_{w,b,\xi,\xi^*} \left(\frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum\nolimits_{i=1}^{n} \left(\xi_i + \xi_i^* \right) \right), \tag{1}$$

$$\text{subject to}: \; y_i - \boldsymbol{w}.\boldsymbol{x_i} - b \leq \epsilon + \xi_i, \tag{2}$$

$$\boldsymbol{w}.\boldsymbol{x_i} - b + y_i \leq \epsilon + \xi_i \leq \epsilon + \xi_i^*, \tag{3}$$

$$\xi_i, \xi_i^* \geq 0 \forall i, \tag{4}$$

where

- $\boldsymbol{w}$ is the weight vector, determining the orientation of the hyperplane. In our context, it represents how each element in the S_{21} feature vector influences the prediction of cable length,
- b is a bias term that adjusts the position of the hyperplane relative to the origin, acting as a baseline or intercept in our cable length prediction,
- ξ_i, and ξ_i^* are two slack variables, accounting for instances where the prediction error exceeds ϵ. They accommodate cables whose S_{21} parameters might be notably challenging for accurate prediction,
- C is a regularization parameter that controls the balance between a low training data error and maintaining a narrow margin, effectively balancing the prediction accuracy with complexity, and
- ϵ sets the acceptable error margin, permitting our predictions to deviate by this amount from the actual value without incurring a penalty; in our problem's context, it is the allowable deviation in meters from the actual cable length.

C. SVM CLASSIFICATION (SVC) FORMULATION

The SVM classification (SVC) model seeks to find a hyperplane that distinctly categorizes the cables, while permitting some flexibility for noisy or challenging data points. The optimization problem can be articulated as:

$$\min_{w,b,\xi} \left(\frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + C \sum\nolimits_{i=1}^{n} \xi_i \right), \tag{5}$$

$$\text{subject to}: \; y_i \left(\boldsymbol{w}.\boldsymbol{x_i} + b \right) \geq 1 - \xi_i, \tag{6}$$

$$\xi_i \geq 0 \forall i, \tag{7}$$

where $\boldsymbol{w}$ and b denote the weight vector and bias, respectively. Specifically, $\boldsymbol{w}$ signifies how each element in the S_{21} feature vector contributes to the cable's classification, whereas b establishes a foundational baseline for this classification process. Accommodating for instances where a cable's S_{21} parameters pose classification challenges or to account for noisy data points, ξ_i are introduced as slack variables. Additionally, the parameter C acts as a balancing factor. It harmonizes the pursuit of a substantial margin—ensuring robust classification—with the imperative to minimize classification inaccuracies, particularly pertinent for cables proximate to the decision hyperplane.

D. KERNEL TRICK

In order to predict the cable lengths based on their S_{21} parameters, an essential consideration is the nature of the relationship between these parameters (predictors) and the actual cable lengths (responses. Support Vector Machines (SVM) employ a mechanism known as the "kernel trick" to capture and represent complex relationships in the data [15]. This technique maps the input data into a higher-dimensional space, thereby facilitating the capture of complex relationships. In this study, we have chosen to employ a polynomial kernel defined as

$$K(u, v) = (u.v + c_0)^p, \tag{8}$$

where c_0 operates as an independent coefficient in the kernel, calibrating the kernel function. On the other hand, the parameter p defines the degree of the polynomial. This notably offers insights into the complexity and depth of the nonlinear relationships the model can detect. By modulating the degree, we can influence the model's capacity to capture complex patterns inherent in the data. Both u and v are feature vectors, and the kernel function is computing a similarity measure between them. For instance, during training, u might be x_i and v might be x_j, two different data points from the training set.

E. DECISION FUNCTION

After solving the optimization problem for SVM, the model makes predictions based on the decision function. This function calculates a weighted sum of kernel evaluations between a new data point and the support vectors from the training set.

The decision function is mathematically expressed as:

$$f(x) = \sum_{i=1}^{n} \alpha_i K(\boldsymbol{x}, \boldsymbol{x_i}) + b, \tag{9}$$

where α_i are the Lagrange multipliers obtained from the dual optimization problem. For classification, $f(x)$ determines the class label (cable length class) based on its sign. For regression, it provides the predicted value (predicted cable length). The Kernel function, $K(\boldsymbol{x}, \boldsymbol{x_i})$, computes the similarity between the input vector $\boldsymbol{x}$ and a training vector $\boldsymbol{x_i}$.

F. PROPOSED SVM ALGORITHM

The following steps elucidate a rigorous SVM-based methodology tailored to calculate $f(x)$ in (9), and hence predict/classify the cables in the testing set:

1. **Data Preparation**: Extract S_{21} parameters—magnitude, phase, and frequency—from the dataset. Each parameter serves as a feature for the SVM regression/Classification model.
2. **Combining Features**: For every cable, concatenate its magnitude and phase data, resulting in a single feature vector.
3. **Data Splitting**: Partition the dataset into training and test sets, typically using 80% of the data for training and 20% for testing.
4. **Feature Scaling**: Normalize the data features to a similar scale, ensuring no particular feature disproportionately influences the model. This step ensures that each feature contributes equally to the model's prediction/classification.
5. **Model Training**: Using the training set, train an SVM regression or classification models with a polynomial kernel. This involves solving the optimization problems in (1)-(4) and (5)-(7), respectively, with the help of tools like GridSearchCV to determine the best hyperparameters.
6. **Hyperparameter Tuning**: Utilize tools like GridSearchCV or RandomizedSearchCV to identify the optimal hyperparameters for the SVM model.
7. **Prediction/Classification**: Apply the trained model to the test set to predict cable lengths.
8. **Performance Evaluation**: Use metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), R^2 (for regression), and Accuracy, Precision, Recall, and F1-Score (for classification) to assess the model's effectiveness.
9. **Optimization**: Based on the results from the performance evaluation, refine and optimize the model further, if necessary. This could involve feature engineering, selecting a different kernel for SVM, re-tuning hyperparameters, or even gathering more data.
10. **Final Prediction/Classification**: Make the final predictions or classifications with the optimized model.

Figure 4 depicts a flowchart illustrating the SVM-based methodology tailored for predicting and classifying cable lengths based on S_{21} parameters.

In our study, we opted for a cable-wise data split. This means all data from a single cable is either used for training or testing, but not both. The reason? Cables can have unique patterns and by using this split method, we ensure the model

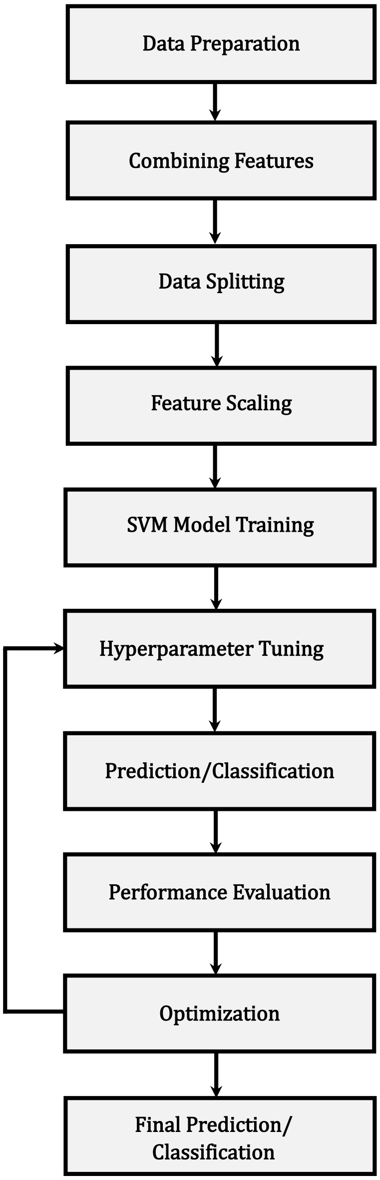

FIGURE 4. SVM-based methodology flowchart.

learns to predict new cables it has not seen before, rather than just memorizing patterns from known cables. This approach tests the model's ability to handle real-world scenarios that might encounter unfamiliar cables.

V. PERFORMANCE MEASURES

In order to quantify the performance of the SVM regression model (SVR) and the SVM classification model (SVC), a suite of metrics has been adopted, as detailed in Table 2. These performance metrics offer a comprehensive evaluation of our classification models, enabling us to discern the model's strengths and areas of improvement. It is noteworthy to mention that in the context of the table, *TP* stands for True Positives, *FP* represents False Positives, P_c indicates the Number of Correct Predictions, and P_t signifies the Total Predictions made.

VI. SIMULATION RESULTS

The proposed SVM Regression Model (SVR) and SVM Classification Model (SVC), detailed in Section IV-F, were applied to the dataset comprising the S_{21} measurements described in Table 1. A polynomial kernel was utilized for both models. This section details the obtained simulation results.

TABLE 3. Performance measures used to assess the SVR and SVC models.

Performance Measure	Definition	Meaning
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - \hat{y}_i\rvert$	Average of absolute errors between predicted and actual values.
Mean Squared Error (MSE)	$\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$	Average of the squares of the errors between predicted and actual values.
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$	The square root of the average of the squared differences between predictions and actual values.
Mean Absolute Percentage Error (MAPE)	$\frac{100}{n}\sum_{i=1}^{n}\frac{\lvert y_i - \hat{y}_i\rvert}{\lvert y_i\rvert}$	Average of absolute percentage errors relative to actual values.
Coefficient of Determination (R^2)	$1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$	Proportion of the variance in the dependent variable that is predictable from the independent variables.
Accuracy	$\frac{P_c}{P_t}$	Proportion of correct classifications out of the total classifications.
Precision	$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}$	Proportion of actual positive identifications out of all positive identifications made.
Recall (Sensitivity)	$\frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$	Proportion of actual positives that were identified correctly.
F1-Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	Harmonic mean of precision and recall. Provides a balance between the two.

A. SVM REGRESSION (SVR) RESULTS

The proposed SVM Regression Model (SVR) was first tested on Group Y (underground cables) and then on Group X (Cables laid above ground). Figures 5 and 6 show the scatter plots that visually compare the predicted vs. actual cable lengths obtained when applying the SVR model to Group Y and Group X, respectively. The diagonal dashed line represents the ideal scenario where predictions are perfect (i.e., each predicted value matches the actual value). The blue dots represent each test cable's actual length against its predicted length.

The proximity of the predictions to the diagonal line, substantiated by Figures 5 and 6, suggests that the SVM

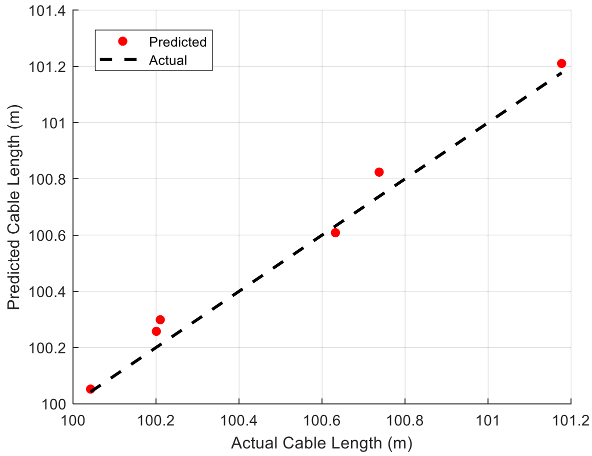

FIGURE 5. Actual versus predicted cable lengths for Group Y using SVR.

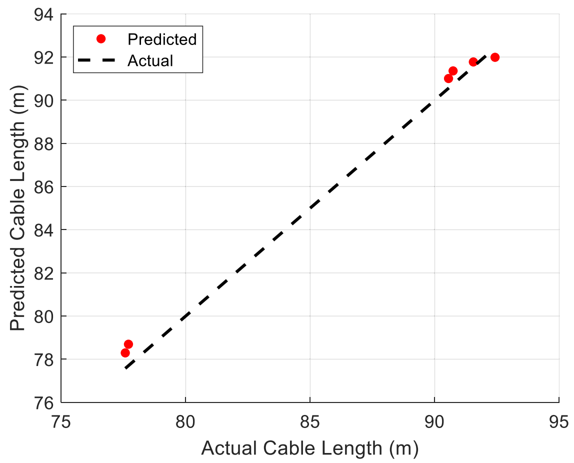

FIGURE 6. Actual versus predicted cable lengths for Group X using SVR.

regression model with a polynomial kernel provides accurate predictions for most test cables in both groups. This further demonstrates the feasibility of employing SVR with S_{21} measurements to predict the physical length of communication cables.

Table 3 provides a detailed evaluation of the SVR model's capability in predicting the lengths of Group Y and Group X cables.

TABLE 4. Performance metrics for the SVR model on Groups Y and X.

Metric	Group Y	Group X
MAE	0.049891	0.572873
MSE	0.003389	0.387743
RMSE	0.058213	0.622690
MAPE	0.049655	0.682689
R^2	0.987828	0.990755

The R^2 value of approximately 0.98 and 0.991 suggests the model has a high level of explanatory power on the test set for Group X and Group Y, respectively. Overall, the model performs well in predicting cable lengths based on their S_{21} measurements for both sets of cables.

The choice of a polynomial kernel further enhances the model's flexibility to capture nonlinear patterns inherent in the dataset. The meticulous hyperparameter tuning process ensured the model's balance between bias and variance, leading to robust predictive performance.

However, like all models, SVM regression has its limitations. The choice of hyperparameters, especially the regularization parameter C and the epsilon-insensitive tube ϵ, plays a crucial role in the model's performance. Their optimal values were contingent on the current dataset and might need recalibration if applied to a different dataset or under varied conditions.

It is worth noting that there are differences in the performance metrics between the two groups. The slightly higher error metrics for Group X may be attributed to the inherent differences in above-ground cables' physical and environmental characteristics compared to buried ones.

B. SVM CLASSIFICATION (SVC) RESULTS

The proposed SVM **Classification** Model (SVC) was similarly tested on cable Groups Y and X. Table 4 comprehensively evaluates the model's performance across varying class granularities.

TABLE 5. Performance metrics of the SVC model for different numbers of classes in Groups Y and X cables.

# of classes	Accuracy (%)		Precision (100%)		Recall (100%)		F1-Score (100%)	
	Y	X	Y	X	Y	X	Y	X
4	90	100.	69	100	75	100	71	100
5	100	83.3	100	66.7	100	75	100	70
6	70	66.7	39	43.3	44	60	41	49.3
7	70	77.8	50	58.3	61	66.7	54	61.1
8	70	66.7	50	36.1	61	50	54	41.1
9	70	55.6	40	27.1	57	37.5	47	30.8
10	70	66.7	50	45.24	62	57.14	54	49.52

Based on the results shown in Table 3, dividing the data into 5 classes yielded the best performance for Group Y. This means that using 5 classes appears to be the most effective approach for predicting cable lengths for Group Y cables based on their S21 measurements. Figure 7 displays the obtained corresponding confusion matrix for the 5-classes case, with a dominant diagonal trend, indicating a high proportion of accurate classifications. On the other hand, using 4 classes gave the best performance for Group X, achieving 100% in accuracy, precision, recall, and F1-Score, as evidenced by the obtained confusion matrix shown in Figure 8. As we increased the number of classes, the performance dropped. This led us to choose 4 classes as optimal.

The heatmap in Figures 7 and 8 visually represents the confusion matrix, showing the number of correctly and incorrectly predicted instances per class. The diagonal of the matrix represents the correct predictions, while other entries indicate misclassifications. Therefore, Figure 7 indicates that the model made no misclassifications on Group Y for the 5-classes case, as all predictions are on the diagonal. Furthermore, the corresponding values for precision, recall, and F1-score for the latter two cases, as shown in Table 3, provide a detailed perspective on the model's performance for each class.

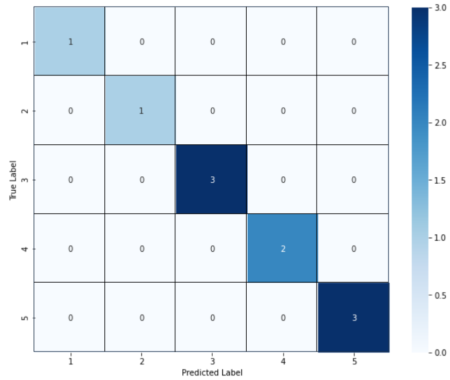

FIGURE 7. Confusion matrix for the SVC model on Group Y.

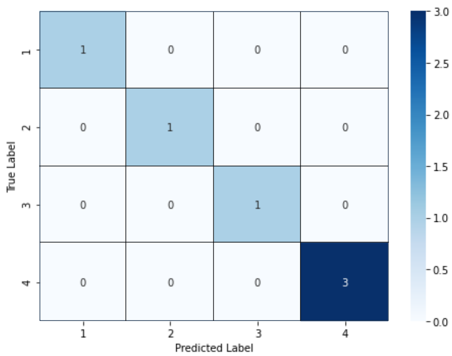

FIGURE 8. Confusion matrix for the SVC model on group X.

Comparing the 5-classes model to the previous 4-classes model for Group Y, introducing an additional class seemingly provides a more granular categorization of cable lengths. A more segmented classification can capture subtle differences in cable lengths more effectively. Additionally, the perfect scores across all performance metrics in the 5-classes model suggest that the division of data into these classes is inherently more representative of the underlying data distribution and the relationships between the S_{21} measurements and the actual cable lengths.

In general, as the number of classes increases, the challenge of distinguishing between neighboring classes also grows, especially if the differences between the classes are subtle. Additionally, there might be a risk of having fewer samples per class, reflected in the declining performance metrics in Table 5.

While more classes give a more detailed classification, balancing granularity and model performance is essential. The model's performance in some classes has dropped compared to the previous models. It is crucial to decide on the number of classes based on both the granularity required and the model's performance in those classes.

VII. CONCLUSION

This work proposes a novel application of Support Vector Machines (SVM) to predict and classify cable lengths based on S_{21} measurements for two distinct groups of communication cables. An SVM regression model (SVR) with a polynomial kernel is built to precisely predict continuous cable lengths derived from the S_{21} measurements (magnitude, phase, and frequency). The model's performance was thoroughly evaluated using several measures, all of which validated the model's robustness and proficiency in predicting the physical length of communication cables understudy. By solving the optimization problem, the SVM ensures a balance between model complexity and prediction accuracy, making it a versatile tool for diverse predictive modeling tasks.

Furthermore, an SVM classification model (SVC) is built to classify cable lengths into distinct classes. Beginning with a simplistic four-class model, we expanded the categorization to encapsulate up to 10 classes. The methodology was underpinned by 1) integration of the magnitude and phase data for holistic representation, 2) Strategically defined class boundaries to ensure distinctive categorization, and 3) implementation of SVM equipped with a polynomial kernel. The model's effectiveness was then assessed using several performance measures, complemented by detailed confusion matrices and classification reports. Based on the simulation results, the SVC model's accuracy exhibited an inversely proportional relationship with the number of classes, revealing the complexities associated with differentiating closely spaced lengths.

The findings from this study emphasize the versatility and robustness of SVM in deciphering cable length attributes from S_{21} measurements. However, the challenge posed by increased classification granularity emphasizes the need for further refinement. Future research may consider investigating more sophisticated machine learning architectures, innovative feature engineering techniques, or employing a richer dataset, especially for lengths with more significant classification challenges.

This manuscript provides a foundational framework for employing SVM in cable length prediction and classification. The insights and results pave the way for continued advancements in this research avenue.

REFERENCES

[1] N. B. Ghazali, D. F. Hashim, F. C. Seman, K. Isa, K. N. Ramli, Z. Z. Abidin, S. M. Mustam, and M. A. Haek, "Cable fault classification in ADSL copper access network using machine learning," *Int. J. Adv. Intell. Informat.*, vol. 7, no. 3, p. 318, Nov. 2021, doi: 10.26555/ijain.v7i3.488.

[2] N. L. Sakun, Z. Z. Abidin, and F. C. Seman, "Cable fault detection in DSL communication system-based on machine learning," *Evol. Electr. Electron. Eng.*, vol. 2, pp. 444–452, Jan. 2021.

[3] H. A. A. Alfeilat, A. B. A. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. B. S. Prasath, "Effects of distance measure choice on K-nearest neighbor classifier performance: A review," *Big Data*, vol. 7, no. 4, pp. 221–248, Dec. 2019, doi: 10.1089/big.2018.0175.

[4] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 51–59, Jan. 2014, doi: 10.1109/TSG.2013.2260421.

[5] K. Hollaus, S. Bauer, M. Leumüller, and C. Türk, "Measurement and modeling of effective cable parameters of unshielded conductors," *COMPEL-Int. J. Comput. Math. Electr. Electron. Eng.*, vol. 41, no. 3, pp. 1041–1051, May 2022.

[6] O. Bader, D. Haddad, A. Y. Kallel, T. Hassine, N. E. B. Amara, and O. Kanoun, "Identification of communication cables based on scattering parameters and a support vector machine algorithm," *IEEE Sensors Lett.*, vol. 5, no. 7, pp. 1–4, Jul. 2021, doi: 10.1109/LSENS.2021.3087539.

[7] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst. Appl.*, vol. 13, no. 4, pp. 18–28, Jul. 1998, doi: 10.1109/5254.708428.

[8] A. Said, S. Hashima, M. M. Fouda, and M. H. Saad, "Deep learning-based fault classification and location for underground power cable of nuclear facilities," *IEEE Access*, vol. 10, pp. 70126–70142, 2022, doi: 10.1109/ACCESS.2022.3187026.

[9] K. Sundaresan, N. Metts, G. White, and A. Cabellos-Aparicio. (2016). *Applications of Machine Learning in Cable Access Networks*. [Online]. Available: https://www.nctatechnicalpapers.com/Paper/2016/2016-applications-of-machine-learning-in-cable-access-networks

[10] S. Mishra, S. C. Swain, T. Biswal, and A. Routray, "Tree based fault classification in underground cable," *Proc. Comput. Sci.*, vol. 218, pp. 524–531, Jan. 2023, doi: 10.1016/j.procs.2023.01.034.

[11] M. L. A. Pozar and M. David, *Microwave Engineering*. Hoboken, NJ, USA: Wiley, 2012.

[12] *R&SZNL Vector Network Analyzer*. Accessed: Oct. 3, 2023. [Online]. Available: https://www.rohde-schwarz.com/us/products/test-and-measurement/network-analyzers/rs-znl-vector-network-analyzer_63493-432704.html

[13] R. A. Nugrahaeni and K. Mutijarsa, "Comparative analysis of machine learning KNN, SVM, and random forests algorithm for facial expression classification," in *Proc. Int. Seminar Appl. Technol. Inf. Commun.*, Aug. 2016, pp. 163–168, doi: 10.1109/ISEMANTIC.2016.7873831.

[14] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of K-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decis. Anal. J.*, vol. 3, Jun. 2022, Art. no. 100071, doi: 10.1016/j.dajour.2022.100071.

[15] L. Le, J. Hao, Y. Xie, and J. Priestley, "Deep kernel," in *Proc. 3rd IEEE/ACM Int. Conf. Big Data Comput., Appl. Technol.*, Dec. 2016, pp. 1–7, doi: 10.1145/3006299.3006312.

MOHAMMAD AL BATAINEH received the B.S. degree (Hons.) in telecommunications engineering from Yarmouk University, Jordan, in 2003, and the M.S. degree and the Ph.D. degree in electrical engineering from the Illinois Institute of Technology (IIT), USA, in 2006 and 2010, respectively. After his academic pursuits, he held noteworthy positions with institutions, including Yarmouk University, where he was promoted to an Associate Professor, in 2018, and roles with the Argonne National Laboratory and MicroSun Technologies. In August 2020, he joined United Arab Emirates University (UAEU), as an Assistant Professor. His research interests include the application of communications, coding theory, and information theory to the interpretation and understanding of information flow in biological systems, particularly gene expression. His additional research avenues encompass machine learning, network information theory, and optimization.

MALIK MOHAMED UMAR received the B.Eng. degree in electronics and communication engineering from Anna University, Chennai, India, in 2009, the M.Tech. degree in applied electronics from PRIST University, India, in 2013, and the Ph.D. degree in systems engineering from the University of Brunei Darussalam, Brunei Darussalam, in 2019. He is currently a Postdoctoral Researcher with the College of Engineering (COE), specifically with the Electrical and Communications Engineering Department, United Arab Emirates University. His research interests include smart sensors, image processing, machine learning, and the Internet of Things (IoT).

AQUIB MOIN received the M.Sc. degree in radio astronomy and space science in Sweden, and the Ph.D. degree in astrophysics in Australia. He was an Associate Professor with the Department of Physics, College of Science, and an Affiliated Scientist with the National Space Science and Technology Centre (NSSTC), United Arab Emirates University (UAEU). He is currently the PI of the UAEU Ground-based Radio Array Facility and the Gamma-Ray Burst (GRB) Research Program. He also serves as the Coordinator for the B.Sc. Space Science Track and the M.Sc. Space Science Program. His research interests include observational radio astronomy, but he is also involved in multidisciplinary projects covering other related areas of space science.

MOUSA I. HUSSEIN (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from West Virginia Tech, Montgomery, WV, USA, in 1985, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Manitoba, Winnipeg, Canada, in 1992 and 1995, respectively. From 1995 to 1997, he was with the Research and Development Group, Integrated Engineering Software Inc., Winnipeg, developing EM-specialized software based on the boundary element method. In 1997, he joined the Faculty of Engineering, Amman University, Amman, Jordan, as an Assistant Professor. He is currently a Professor with the Department of Electrical and Communication Engineering, United Arab Emirates University. He has supervised several M.Sc. and Ph.D. students. He has over 120 publications in international journals and conferences. His current research interests include computational electromagnetics, electromagnetic scattering, antenna analysis and design, metamaterial and applications, material/bio-material characterization, and sensor design for bio applications.

MAHMOUD AL AHMAD (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from Birzeit University, in 1999, and the M.Sc. and Ph.D. degrees in microwave engineering from Technische Universität München, in 2002 and 2006, respectively. He is currently a Full Professor of electrical and communication engineering with United Arab Emirates University. With over 20 years of research experience in electronic materials and device fabrication across academia, national laboratories, and industry, he has successfully managed research projects with budgets reaching up to U.S. $1 million. He has contributed to approximately 160 published papers and delivered over 50 presentations at international conferences, some of which were by invitation. Moreover, he holds more than 30 U.S. patents. His research interests include innovative areas, including the design and development of self-powered, low-powered nano-based electronic devices and systems, along with applied electromagnetic research for biomedical applications. He is a Senior Member of IEEE with a rich academic and research background.