Design of interdigital filters for millimeter-wave applications using artificial neural networks

Luis A. S. Tapia, Luis H. F. Oraggio, Carla R. Martins, Mirian P. dos Santos, José A. Oliveira, Ivan Aldaya, and Rafael A. Penchel

***Abstract*— This work presents a purpose to accelerate the design process of mmWaves filters for 5G applications. To aim this objective, this work makes use of a technique of an artificial neural network called the multilayer perceptron. The model has 4 outputs and 6 inputs, where 2 outputs have a strong correlation and the other two have a weak correlation. By using 1000 samples, the model achieves an accuracy value of 0.927 for the outputs with strong correlation, but despite the good result, the predicted model still needs improvement, since $|S_{11}|$ had a not-so-nice performance.**

***Keywords*— Interdigital filter, artificial neural networks, millimeter-wave.**

I. Introduction

5G technology presents a significant revenue opportunity for consumer services, as demonstrated in [1]. Globally, the total mobile data traffic is projected to reach 325 exabytes (EB) per month by 2028, which is approximately four times the amount recorded at the end of 2022 when mobile data traffic stood at 90 EB, as reported in [2]. To shed more light on how 5G can contribute to revenue, a survey conducted by [3] involving 3,000 consumers revealed that over half of the respondents were willing to pay more for the benefits of 5G. This demonstrates that the cost is not a prohibitive factor for adoption. This is due to the fact that 5G enhances user experiences for applications demanding high data traffic, security, and low latency.

While the millimeter wave (mmWave) range spans 30 GHz to 300 GHz, companies are terming the higher frequencies ($\geq$ 24 GHz) within this range as '5G mmWaves'. This nomenclature arises from their proximity to the mmWave range and similar benefits, such as higher data rates. However, these frequencies suffer from path loss, limiting their range and making them suitable for short distances, which enhances security and privacy. Due to the high frequency, circuit components must be smaller than the wavelength, necessitating a distributed circuit-element approach[4].

Luis A. S. Tapia, FESJ/UNESP, São João da Boa Vista - SP, e-mail: luis.tapia@unesp.br; Luis H. F. Oraggio, FESJ/UNESP, São João da Boa Vista - SP, e-mail: lh.oraggio@unesp.br; Carla R. Martins, FESJ/UNESP, São João da Boa Vista - SP, e-mail: c.martins@unesp.br; Mirian P. dos Santos, FESJ/UNESP, São João da Boa Vista - SP, e-mail: mirian.santos@unesp.br; Jose Augusto de Oliveira, FESJ/UNESP, São João da Boa Vista - SP, e-mail: jose.a.oliveira@unesp.br; Ivan Aldaya, FESJ/UNESP, São João da Boa Vista - SP, e-mail: ivan.aldaya@unesp.br; Rafael A. Penchel, FESJ/UNESP, São João da Boa Vista - SP, e-mail: rafael.penchel@unesp.br. This work was supported in part by the Brazilian Agency Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) under Grant 2020/09889-4, in part by the Brazilian Agency Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq) under Grants 313378/2021-5 and 409146/2021-8, and in part by the Brazilian Agency FINEP under Grant 0527/18.

Specialized software like Advanced Design System (ADS) and Ansys EM is crucial for designing mmWave components. ADS uses approximate methods for faster results, while Ansys EM employs full-wave electromagnetic analysis for more reliable data. Simulating component responses efficiently remains a challenge, with machine learning techniques offering a potential solution.

Machine learning is characterized by its capacity to learn from the environment. It's a computational algorithm branch addressing various problems, including finance, entertainment, medical applications, and more. Machine learning adapts its architecture through generations, producing superior results without 'hard code'. Artificial Neural Networks (ANNs), a subset of machine learning, simulate biological learning through interconnected neurons. These neurons compute functions using input data and weights, enabling learning when connected with other neurons [5], [6], [7].

This study aims to create a 5G mmWave interdigital filter for the 26GHz to 28GHz range, utilizing the ROGERS RT/duroid 6006 substrate. Initial results will be obtained via Ansys EM analysis, followed by utilizing an artificial neural network to determine the filter's dimensions based on S parameters.

II. Filter modeling using neural network

In this section, will be explained briefly about neural network and the interdigital filter design process.

A. Neural network

Neural networks, inspired by the natural learning mechanisms, employ artificial counterparts called perceptrons [8]. Within living organisms, neurons form the foundation of neural networks, comprised of three essential components: the cell body, dendrites, and axon. Dendrites act as signal receptors from nearby neurons, while axons convey these signals. The point of connection between an axon terminal and a dendrite is referred to as a synapse (refer to Figure 1 for visualization).

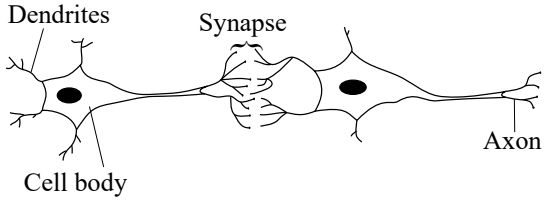
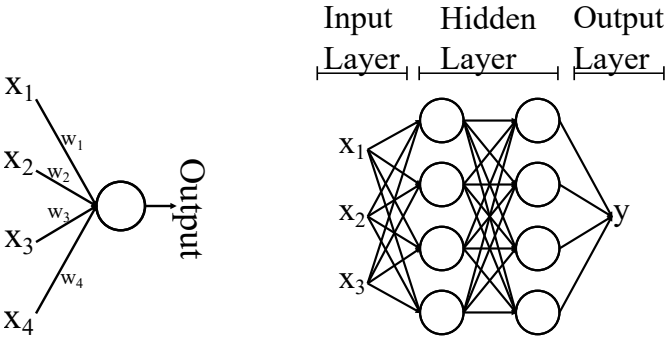


Fig. 1: Biological neuron: three main parts. Based in [7]

The artificial representation of neural networks was introduced with the development of perceptrons, a type of

artificial neuron. Perceptrons receive binary inputs or features (x_1, x_2, ..., x_n) to generate a single binary output. Each input is assigned a weight (w_1, w_2, ..., w_n) that determines its importance for the desired output. In an artificial neural network with multiple layers of perceptrons, the output of one perceptron can serve as the input for another, governed by the assigned weights. This mechanism mimics the role of synaptic connection strengths in biological organisms.

The predicted output of a perceptron can take values of 0 or 1, determined by a weighted sum ($\sum_j w_j x_j$) and a threshold value. This process is illustrated in Figure 2a. In some systems, additional components called biases are introduced to enhance prediction accuracy. Biases help the model better adapt to various situations.



(a) Simple perceptron representation

(b) Multilayer perceptron

Fig. 2: Perceptron definition

The learning process in artificial neural networks (ANN) revolves around altering the weights connecting neurons. This process is stimulated by training data, encompassing pairs of input and corresponding output. In biological organisms, incorrect reasoning triggers an unpleasant response; similarly, in neural networks, the training data evaluates the weights based on the discrepancy between predicted and actual output, striving to minimize this discrepancy. The mechanism to minimize this disparity is termed the loss function. The process of comparing predicted and target outputs and adjusting weights accordingly is called backpropagation [7].

To evaluate the predictive precision of a developed model, the mean squared error (MSE) is a commonly employed metric ($\frac{1}{n}\sum(y_i - f(x_i))^2$), wherein 'n' represents the total number of observations, y_i signifies the target output, and $f(x_i)$ indicates the predicted value. This measure is typically computed on the test dataset. The construction of an Artificial Neural Network (ANN) model entails partitioning the dataset into two subsets: the training data, where the learning process unfolds, and the testing data, utilized to assess the model's accuracy.

Nevertheless, while this approach offers insights into model performance, it can falter if the training data exhibits an uneven distribution of specific cases. This can lead to skewed learning and hinder the model's ability to generalize. To counteract this challenge, alternative techniques like cross-validation leverage data resampling methods. A prominent approach is k-fold cross-validation, where the dataset is divided into 'k' subsets or folds. In each iteration, one fold is set aside for testing while the remaining folds contribute to training. This procedure is repeated 'k' times, and the average MSE across iterations provides the comprehensive score—a method known as k-fold cross-validation [7] [9].

Activation functions play a pivotal role in the transformation of inputs to outputs within neural networks. This process begins by calculating the sum of weighted inputs, which is then passed through an activation function. The importance of this function lies in its ability to shape prediction accuracy. Without it, the output remains linear, constraining the network's capacity to handle intricate data mappings. Activation functions come in two types: linear and non-linear, with the latter being more widely embraced due to their effectiveness in accommodating real-world non-linear error patterns [10].

Figure 3 showcases a range of activation functions, each chosen based on specific criteria. In the context of multi-layered neural networks, non-linear functions such as step, sigmoid, or tanh find preference. When prioritizing genuine prediction goals and interpretability, the identity function emerges as a suitable choice. For binary class label predictions, the step function demonstrates its efficacy. Meanwhile, estimating binary class probabilities calls for the sigmoid function, with the symmetric tanh function often surpassing sigmoid due to its symmetry. The widely adopted Rectified Linear Unit (ReLU) function owes its popularity to its ability to selectively activate neurons [7], [10].

The identity activation function retains the input as the output ($f(x) = x$). However, its constant gradient inhibits effective error reduction through gradient-based techniques. On the other hand, the sigmoid activation function ($f(x) = \frac{1}{1+e^{-x}}$), which maps inputs into the 0 to 1 range, introduces training challenges due to its asymmetric and constant-sign output. In contrast, the tanh activation function ($f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$) yields an output range of -1 to 1, efficiently encompassing the spectrum of output values. This is further augmented by its zero-centered nature, enhancing mapping capabilities [10].

The ReLU activation function's distinctive feature is its selective neuron activation, rendering neurons with linear transformation outputs of 0 inactive. ReLU's derivative function effectively facilitates backpropagation ($f(x) = max(0, x)$) [10].

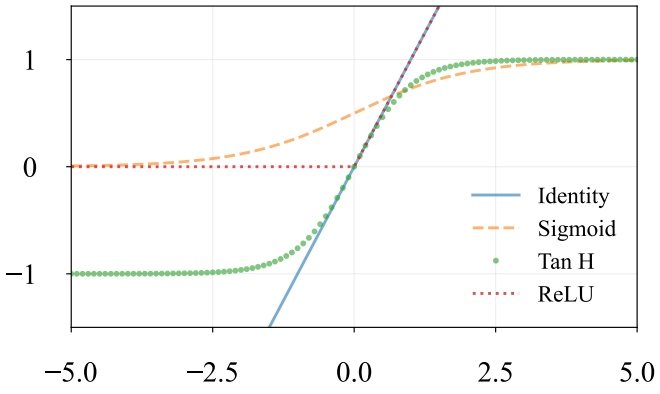


Fig. 3: Active functions.

Moreover, for tackling more intricate problems, it is advi-

sable to employ a network of neurons, as depicted in Figure 2b. In this architecture, the decisions made in the initial layer are simplistic. However, as additional layers are introduced leading to the output, the decisions become progressively complex. The layers situated between the input and output are commonly termed 'hidden layers,' as the computations occurring in these layers remain concealed from users. This hidden layer can incorporate bias adjustments to enhance predictions. It's worth noting that a neural network can encompass 'n' perceptrons within each layer, thereby accommodating multiple outputs [11], [7].

B. Interdigital filter

An essential component for wireless systems is the filter, used to confine signals within specific frequency ranges [12]. The filter design process involves considering various parameters, such as the filter type, response, and order. Filters can take various forms, including low-pass filters (allowing signals below the cutoff frequency, f_c), high-pass filters (permitting signals above f_c), pass-band filters (operating within the f_{c1} and f_{c2} range), and band-stop filters (acting as the opposite of a band-pass filter). Furthermore, filter responses can exhibit characteristics like Butterworth responses (maximally flat in the passband) or Chebyshev responses (with passband ripple and steeper transition bands). The filter order also plays a crucial role, influencing complexity and transition band steepness; higher orders result in greater complexity but enhanced selectivity [13].

One prevalent technique for managing pass and stop band characteristics, along with phase behavior, is the insertion loss method. This approach implies constructing a π network using inductors, capacitors, and resistors. Figure 4 showcases a 5th-order π network as an illustrative example. The values of capacitance and inductance are showed within the figure."

$$L'n = \frac{\Delta Z_0}{\omega_0 g_N} \quad \text{and} \quad C'n = \frac{g_N}{\omega_0 \Delta Z_0} \tag{1}$$

for parallel elements and

$$L'n = \frac{g_N Z_0}{\Delta \omega_0} \quad \text{and} \quad C'n = \frac{\Delta}{\omega_0 g_N Z_0} \tag{2}$$

for series elements, where g_n is the normalized parameter, that changes with order and kind of response of filter, Z_0 is characteristic impedance of the line (usually $Z_0 = 50\Omega$) and $\Delta = \frac{w_2 - w_1}{w_c}$ (fractional bandwidth).

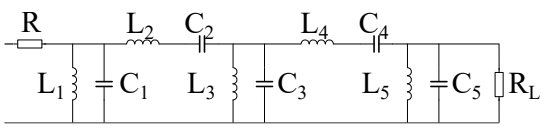


Fig. 4: 5-order π network

Utilizing distributed element circuits offers a viable approach to achieving outcomes similar to those obtained with lumped element circuits. This makes them a valuable alternative for designing electronic components intended for higher frequencies [4]. To achieve comparable performance, distributed element circuits employ $\lambda_g/4$ resonators, where $\lambda_g = \lambda/\sqrt{\epsilon_r}$ [14].

The interdigital topology is an extension of the hairpin design, employing an array of resonators with an electrical length of $\lambda_g/4$. In contrast to the virtual short circuit characteristic of a hairpin filter, the interdigital filter incorporates a real ground. This distinction arises from the placement of a through-substrate via (TSV) at one edge of each resonator and an open circuit at the opposite end. This arrangement is necessary to achieve impedance and reactance equivalence with a lumped circuit [12]

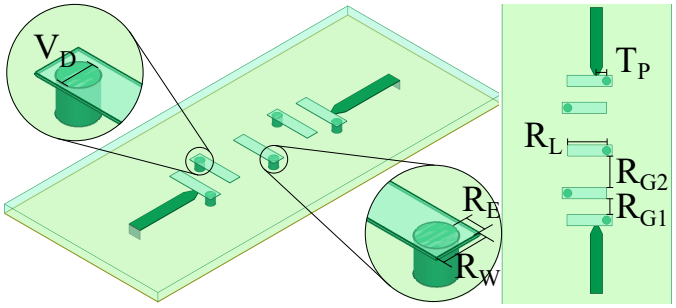


Fig. 5: Interdigital filter structure

The design of a 5th-order interdigital filter is depicted in Figure 5. In this design, the resonator's dimensions are determined by the values of R_L for length and R_W for width. This structure features symmetry concerning the central resonator, resulting in only two distinct gap sizes between resonators, namely R_{G1} and R_{G2}. To achieve an enhanced response, a supplementary length denoted as R_E is introduced for odd resonators. Furthermore, the electrical distance (T_P in radians) between the tapping point and the through-substrate via (TSV) located at the short circuit is a crucial parameter. The initial design equations are as follows:

$$\theta = \frac{\pi}{2}\left(1 - \frac{\Delta f}{2 f_o}\right) \quad \text{and} \quad Y = \frac{Y_o}{\tan\theta} \tag{3}$$

$$Y = \frac{Y_o}{\tan\theta} \tag{4}$$

where Δf is the pass-band, f_o the central frequency and $Y_o = 1/Z_o$ the input admittance ($Z_o = 50\Omega$).

III. Results

The initial phase of designing a 5th-order Butterworth filter entailed the creation of an ideal response, conceptualized using a π network and represented through equations (1) and (2). The circuit's layout followed the configuration illustrated in Figure 4, with the specific values of inductance, capacitance, and resistance showed in Table I.

Subsequent to this, Advanced Design System software was employed to calculate the dimensions of distributed elements, employing equations (3) and (4). This calculation step was then complemented by further dimension refinement using the Nuhertz Designs System. After this, with Ansys EM, the final

design of the filter was executed. This design strategy incorporated a substrate thickness of 0.254 mm and a conductor thickness of $0.18\,\mu m$.

TABLE I: Values of lumped-element pass-band filter

Element	Abbreviation	Value
Input impedance	R	50 Ω
Load impedance	R_L	50 Ω
Capacitor 1	C_1	0.983 pF
Capacitor 2	C_2	5.405 fF
Capacitor 3	C_3	3.183 pF
Capacitor 4	C_4	5.405 fF
Capacitor 5	C_5	0.983 pF
Inductor 1	L_1	35.37 pH
Inductor 2	L_2	6.438 nH
Inductor 3	L_3	10.93 pH
Inductor 4	L_4	6.438 nH
Inductor 5	L_5	35.37 pH

To initiate the process, a prototype of the filter was developed using Nuhertz Filter Solutions. The S parameters for this prototype are illustrated in Figure 8, and the corresponding dimensions are detailed in Table II under the 'Initial Design' column. Subsequently, a manual optimization was conducted to determine an optimal solution, as depicted in Figure 5. In order to discern the variables exerting the most influence on the S parameters, a parametric analysis was performed, taking into account R_L, R_W, R_E, R_{G1}, R_{G2}, and T_P.

Through this analysis, it became evident that the variables carrying the greatest impact were R_{G1}, R_{G2}, R_E, and T_P. As consequence, a dataset comprising 1000 samples was generated. Each sample involved variations of R_{G1}, R_{G2}, R_E, and T_P within a range of $\pm$ 20 percent of the manually optimized result.

TABLE II: Structure and dimensions: initial and final filter

Item	Dimensions	Size [μm] Initial design	Size [μm] Final design
Resonator's width	R_W	409	364
Resonator's length	R_L	1124	1150
Resonator extension length	R_E	12.65	20
Gap 1 between resonators	R_{G1}	539.2	381
Gap 2 between resonators	R_{G2}	660.8	680
Tap-point electrical length	T_P	197.4	300
Through-substrate-vias diameter	V_D	125	125
Feed-line length	F_L	300	300
Feed-line width	F_W	67	67

A correlation analysis was performed using inputs from treated raw data. For this problem, the input parameters are S_{11} bandwidth, which is the range of $|S_{11}|$ that are -10 dB in relation to the maximum value of it; S_{11} average, which corresponds average of $|S_{11}|$ that are inside the previously input; $S_{11}F_c$, which corresponds to the F_c of S_{11} bandwidth; S_{21} bandwidth, which is the range of $|S_{21}|$ that are in relation of the fallen of 3 dB to the maximum value of it; S_{21} average, which corresponds to the average of $|S_{21}|$ that are range of the previously input; $S_{21}F_c$, which corresponds to the F_c of S_{21} bandwidth. In addition, the outputs are R_{G1}, R_{G2}, R_E, and T_P.

The correlation analysis highlighted that only R_{G1} and R_{G2} exhibited a strong correlation with the input parameters. For instance, an increase in R_{G1} corresponded to a reduction in bandwidth for S_{11}, S_{21}, and a decrease in the average value of S_{21}. The same trend was observed for R_{G2}. Conversely, weak correlations were identified between R_E and T_P with the input parameters. Consequently, two models of Artificial Neural Networks (ANNs) using Multilayer Perceptrons (MLPs) were constructed: one model incorporated the more strongly correlated parameters, while the other included the weaker ones.

The optimal activation function was determined through a comparative analysis, utilizing 100 neurons and a single hidden layer. Additional layers were tested but resulted in inferior performance. The learning rate was held constant at 0.01 for this assessment. In the first model, where parameters exhibited a negative correlation with the output, the tanh activation function emerged as the most effective choice. It displayed an elevated accuracy (with a score of 0.927), surpassing the performance of identity (0.896), ReLU (0.909), and sigmoid (0.924) functions. This heightened accuracy could be attributed to its inverse proportionality to the parameters. The loss function, visualized in Figure 6, indicated higher values for data output. Elevated accuracy and loss function values suggest that the model makes infrequent yet impactful errors.

Further examination identified the optimal configuration as comprising 200 neurons in a single hidden layer

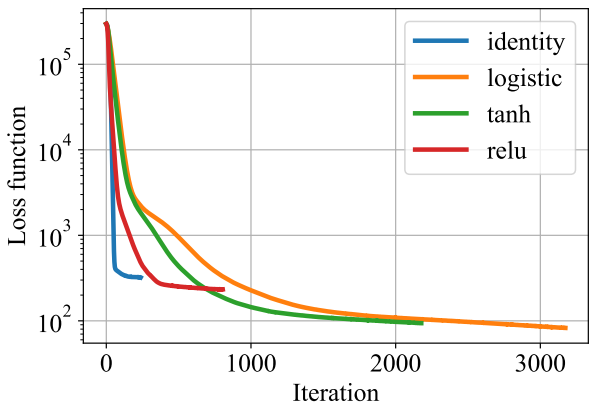
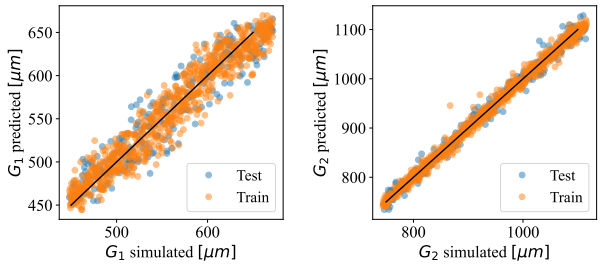


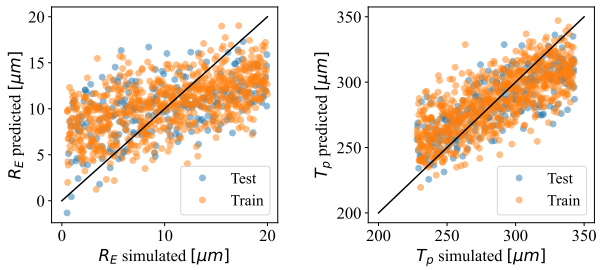
Fig. 6: Parametric analysis of activation functions for G_1 and G_2.

In a similar manner, a study was conducted to determine the optimal model for the R_E and T_p parameters. In this case, due to their weaker correlations, the choice of activation function was guided by computational resource efficiency. The resulting accuracy was relatively low, and the loss function was high. A comparison of the four parameter performances can be seen in Figure 7. The first group of parameters exhibited higher accuracy, whereas the second group performed less effectively.

Moreover, Figure 8 illustrates the responses of $|S_{11}|$ and $|S_{21}|$ for the lumped filter design, Nuhertz solution, and MLP solution based solely on the predictions of the first parameter group. Given the low correlation of T_P and R_E with the



(a) Predicted X Trained data: G_1 and G_2.



(b) Predicted X Trained data: R_E and T_P

Fig. 7: Predicted X Trained data

input parameters, their impact remains minimal. The MLP prediction outperforms the Nuhertz solution, yet it falls short of the ideal outcome. The predictions were generated with S_{11} and S_{21} bandwidths set at 2 GHz, f_c of S_{11} and S_{21} at 27 GHz, and average $|S_{11}|$ and $|S_{21}|$ at -100 dB and -3 dB, respectively. It's worth noting that the frequency range where $|S_{11}| \leq -10dB$ extends from 25.992GHz to 28.1GHz, and the average $|S_{22}|$ is 26.98dB, indicating the ANN's effective evaluation of other parameters. The presence of a peak within the S_{11} bandwidth occurs because this input parameter only considers the frequencies where $|S_{11}| \leq -10\ dB$, without considering behaviors within this interval

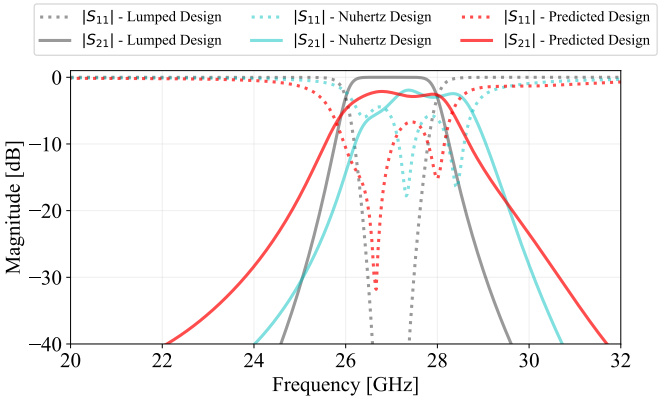


Fig. 8: S parameters of filter design.

IV. Conclusions

In brief, the purpose of this project was to use the ANN to accelerate the design process of a filter that works for 5G applications, in specific, a 5-order interdigital filter that has a Butterworth response and operates at 26 GHz to 28 GHz in ROGERS RT/duroid 6006. Many targets were achieved, but the model needs improvements, like the bandwidth of return loss or the usage of a higher order filter to have more parameters that have a strong correlation with S parameters.

References

[1] Ericsson, "Uncovering the revenue opportunity of 5g," Aug 2020. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/consumerlab/reports/harnessing-the-5g-consumer-potential

[2] "Mobile data traffic forecast – mobility report," Jan 2023. [Online]. Available: https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast

[3] "5g report: The value of 5g services and the opportunity for csps." [Online]. Available: https://www.nokia.com/thought-leadership/research/5g-consumer-market-research/

[4] J. E. G. Lé, M. Ouvrier-Buffet, L. G. Gomes, R. A. Penchel, A. L. C. Serrano, and K. G. P. Rehder, "Integrated Antennas on MnM Interposer for the 60 GHz Band," *Journal of Microwaves, Optoelectronics and Electromagnetic Applications*, vol. 21, pp. 184–193, Mar. 2022.

[5] I. El Naqa and M. J. Murphy, "What is machine learning?" *Machine Learning in Radiation Oncology*, p. 3–11, 2015.

[6] I. Aldaya, L. Cardinal, J. Menzinger, R. A. Penchel, M. O. Santos, M. Abbade, J. A. Oliveira, M. Santos, and G. G. Sánchez, "Integration of a* and k-means clustering for star network design in environments considering obstacles," *Anais do XL Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 2022.

[7] C. C. AGGARWAL, *Neural networks and deep learning: A textbook*. SPRINGER INTERNATIONAL PU, 2018.

[8] P. Vadapalli, "Biological neural network: Importance, components amp; comparison," Oct 2022. [Online]. Available: https://www.upgrad.com/blog/biological-neural-network/

[9] D. Berrar, "Cross-validation," *Encyclopedia of Bioinformatics and Computational Biology*, p. 542–545, 2019.

[10] S. Sharma, S. Sharma, and A. Athaiya, "Activation functions in neural networks," *International Journal of Engineering Applied Sciences and Technology*, vol. 04, no. 12, p. 310–316, 2020.

[11] M. A. Nielsen, *Neural networks and deep learning*. Determination Press, 2015.

[12] J.-S. Hong, *Microstrip filters for RF/Microwave Applications*. Wiley, 2011.

[13] D. M. Pozar, *Microwave engineering*, 4th ed. Wiley, 2012.

[14] R. Karim, A. Iftikhar, and R. Ramzan, "Performance-issues-mitigation-techniques for on-chip-antennas – recent developments in rf, mm-wave, and thz bands with future directions," *IEEE Access*, vol. 8, pp. 219 577–219 610, 2020.