Received 8 September 2020; revised 24 October 2020; accepted 1 November 2020. Date of publication 3 November 2020;
date of current version 25 February 2021. The review of this article was arranged by Editor J. Kumar.

Digital Object Identifier 10.1109/JEDS.2020.3035628

Large-Signal Modeling of GaN HEMTs Using Hybrid GA-ANN, PSO-SVR, and GPR-Based Approaches

ANWAR JARNDAL [1] (Senior Member, IEEE), SADDAM HUSAIN [1,2], MOHAMMAD HASHMI [2] (Senior Member, IEEE), AND FADHEL M. GHANNOUCHI [3] (Fellow, IEEE)

1 Electrical Engineering Department, University of Sharjah, Sharjah, UAE
2 Electrical and Computer Engineering Department, Nazarbayev University, Nur-Sultan 010000, Kazakhstan
3 iRadio Lab, Department of Electrical and Computer Engineering, University of Calgary, Calgary, AB T2N 1N4, Canada

CORRESPONDING AUTHOR: A. JARNDAL (e-mail: ajarndal@sharjah.ac.ae)

This work was supported by the University of Sharjah, Sharjah, UAE.

ABSTRACT This article presents an extensive study and demonstration of efficient electrothermal large-signal GaN HEMT modeling approaches based on combined techniques of Genetic Algorithm (GA) with Artificial Neural Networks (ANN), and Particle Swarm optimization (PSO) with Support Vector Regression (SVR). Another promising Gaussian Process Regression (GPR) based large-signal modeling approach is also explored and presented. The GA-ANN addresses the typical problem of local minima associated with the backpropagation (BP) based ANN. The GA successfully aids in the determination of optimal initial values for BP-ANN and enables it to find a unique optimal solution after subsequent of iterations with higher rate of convergence. This is also achieved using PSO-SVR with lower optimization variables. The developed modeling techniques are demonstrated and used to simulate the gate and drain currents of a 2-mm GaN device. All the models are relatively simple, practical, and easy to implement. The gate and drain currents models are embedded in an equivalent large-signal circuit's model and built in Advanced Design System (ADS) software. The implemented model is validated by large-signal measurements and very good fitting results have been obtained. The model also showed an accurate simulation for a nonlinear power amplifier with very good computational speed and convergence.

INDEX TERMS ANN modeling, GaN HEMT, GPR modeling, large-signal modeling, SVR modeling.

I. INTRODUCTION

The power amplifiers employed in broadcasting and communication transmitter applications are high power and are intrinsically non-linear [1]–[7]. Therefore, in these applications, the GaN High Electron Mobility Transistor (HEMT) is becoming an optimal device [8]. This is owing to its ability to provide high output power at high frequency with excellent efficiency to meet the requirements of advanced broadcasting and communication systems [9]–[11]. The device has also higher gain with very good noise characteristics and is therefore a very good candidate for the design of low noise amplifiers and integrated GaN based transceivers [12], [13]. Overall, the reliability of power amplifiers in wireless and broadcasting transmitters depends on the accuracy of the employed GaN HEMT device large-signal models [14]. The model in essence should consider the parasitic effects under high frequency and self-heating under high power derive of operation. Another important effect is the surface and buffer trapping, which reduce the RF power of the transistor amplifiers and increase the memory effects. Both self-heating and trapping are frequency dependent phenomena and represent the source of the well-known memory effects [15]. These issues have strong impact on the GaN HEMT device performance and must be taken into account in their corresponding large-signal models [8].

A number of papers have been reported in literature addressing large-signal modeling of GaN HEMT [16]–[30]. Out of these, some of them rely on robust small-signal modeling approach before the eventual large-signal model developments. In general, the commonly used modeling techniques include table-based, analytical, physics-based, and artificial neural networks (ANN) based. The table-based modeling [16], [17] has lower cost in terms of the development and implementation but with lower speed and rate of convergence. This is due to its discrete nature and therefore leads to its poor performance at high frequency of operation. The analytical modeling, on the other hand, relies on closed-form formulations [18], [19]. The continuous nature of this technique makes it faster with higher rate of convergence and better prediction capability. The drawback of this approach is it requires relatively longer time and substantial efforts for the optimization of the fitting parameters of the model. The other conventional physics-based modeling technique provides more insight about the device's physics and could be an optimal tool for design-oriented technology development [20], [21]. However, this technique requires extra effort to collect technology-based information of the device and also suffers from lower speed of simulation and thus is not appropriate for the design of relatively complicated or multi-cell transistor based circuits. The last few years has seen emergence of learning-based models utilizing tools such as ANN [22]–[30]. This technique is promising as it is simple to develop and offers excellent trade-off between accuracy and simulation time. The black box nature of this technique reduces the cost of searching for proper formula and computation of the model's parameters. It exhibits higher rate of convergence when compared to the analytical modeling and its prediction capability can also be improved by choosing suitable model topology and activation function. It is well known that the ANN based model "learns" the relationship between the input voltages and output current from the measured I-V data, and then efficiently predicts the current value for any input voltage [30]. During training, the ANN calculates the resulting output current at certain input voltages and compares it with the measured current to estimate the error, which is then propagated back through the system to adjust the weights for best fitting. This represents the typical procedure of the widely used Back Propagation (BP) ANN.

The main limitation of the back propagation (BP), as a gradient method, is its higher sensitivity to the initial guess as the solution could get stuck in local minima [31]. To overcome this problem, more effort is needed to find proper initial guess (close to the global minimum), tune the model topology, modify the objective function, or change the activation function [32]–[33]. The local minima issue becomes more obvious in a non-linear problem of larger scale ANN model such as IV device's modeling. Herein, the training processes need to be re-initiated many times to find the best fitting. Also increasing the order of the ANN to improve the model fitting complicates the model implementation in CAD software and affect the convergence of simulation. This aspect can be addressed by utilizing global optimization techniques such as genetic algorithm (GA) and particle swarm optimization (PSO) as they have been found to be very good alternatives to train neural networks [34], [35].

In this article, an efficient and simple GA augmented ANN based large-signal modeling of GaN HEMT is developed and reported. Furthermore, there have been reports of Support Vector Regression (SVR) and PSO augmented SVR based techniques to address various behavioral modeling issues [36]–[41]. In this article, SVR technique is revisited and adapted for large-signal modeling of GaN HEMT. To further improve the performance, a PSO augmented SVR based modelling technique is developed. Finally, Gaussian

Process Regression (GPR) technique [42]–[44] is also exploited to model the drain and gate currents. The Bayesian approach is considered herein to model the regression-based problem a non-parametric modeling process. Unlike ANN and SVR, the GPR determines the probability distribution over all possible admissible functions that fits the data, which make it more robust against measurement errors and outlier measured data. In brief, the main contributions of this article are: (i) development of improved GA augmented ANN and PSO augmented SVR based GaN HEMT modeling techniques, and (ii) demonstration of GPR as a new promising modeling technique for GaN HEMT for the first time.

The proposed modeling technique is a hybrid of "black box" and "equivalent circuits" approaches. The implemented equivalent circuit is physical relevant and could be used to predict some physics related behavior such as structure or technology induced parasitic effects. Also, the addition of some parameter in the model such as thermal factor and RC circuit could be used to predict the device performance under different conditions of trapping and self-heating. In general, the developed modeling approach could provide a very good alternative for both device and circuit designers. The equivalent circuit make it compatible with the device structure and could provide relevant insights about the device physics. On the other side the proposed black box approach could be applied efficiently to simulate intrinsic nonlinear elements with higher rate of convergence, which is needed for complicated circuit and sub-system design.

This article begins by introducing the large-signal equivalent circuit model in Section II and then presents the developed approaches in Sections III, IV and V. The three approaches are validated in Section VI. The transistor model implementation and validation are presented in Section VII and the conclusion is drawn in Section VIII.

II. EQUIVALENT CIRCUIT MODEL

Figure 1 shows the implemented large-signal equivalent circuit model. The dotted area encompasses the intrinsic part of the transistor including intrinsic gate current and capacitance in addition to the drain-source channel current. The self-heating effect and its associated dispersion is characterized by the electro-thermal model in the sub-circuit. Further

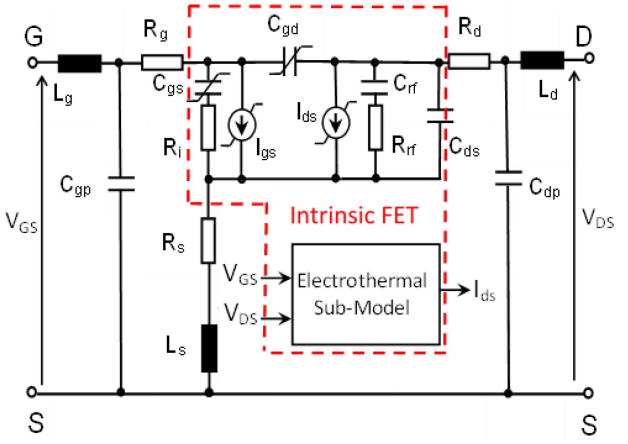

FIGURE 1. Equivalent circuit based electro-thermal large-signal model for GaN HEMTs.

details about this model are provided in the following sections. The trapping effect is described by the $R_{rf}C_{rf}$ circuit in the drain side to synthesize the buffer trapping effects, which has stronger impact of the RF characteristic of the device [17]. The extrinsic part of the model includes lumped elements to represent the resistance, the inductance, and the capacitance parasitic effects. These elements are directly extracted from cold S-parameters measurement for the investigated device [17]. The intrinsic capacitances are extracted from the de-embedded bias-dependent S-parameters. The respective variations with V_{GS} and V_{DS} are mimicked and modelled by simple ANN models as explained in the next sections.

III. GA AUGMENTED ANN BASED MODEL

The proposed architecture consists of two-hidden layers with 4 nodes in each layer as shown in Fig. 2. In the proposed model of drain current there are in total 28 weights and 9 biases. The drain current I_{DS} is modeled using the GA augmented ANN based model, where $I_{DS,\,DC}$ is governed by (1).

$$I_{DS,DC} = w_b^3 + \sum_{k=1}^{4} w_k \tanh \times \left(w_{kb}^2 + \sum_{j=1}^{4} w_{kj} tanh\left(w_{1j} V_{GS} + w_{2j} V_{DS} + w_{bj}^1 \right) \right) \quad (1)$$

where V_{GS}, and V_{DS} are extrinsic gate and drain voltages, respectively. The terms w_{1j}, and w_{2j} are input weights, w_{kj} are the intermediate weights (between two hidden layers). Furthermore, w_{bj}^1, w_{kb}^2 and w_b^3 are the input-layer, hidden-layer and output-layer biases, respectively. The model uses the non-linear $\psi = tanh(.)$ activation or threshold function at each layer. It is also imperative to point out that particular nodes utilize (2) to learn the model parameters. If the output of one neuron is given by k, inputs to the neuron are t_1, $t_2, \ldots, t_n$, weights corresponding to each connection from inputs are $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_n$ and bias as p, then the equation of one neuron model can be formulated by (2).

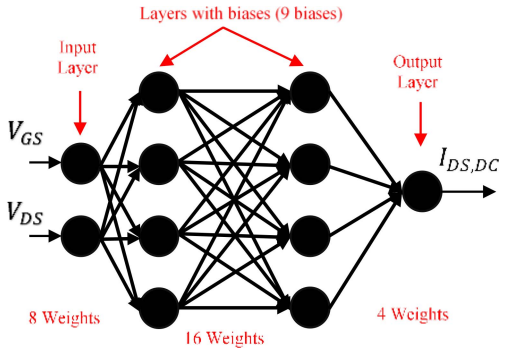

FIGURE 2. The proposed ANN architecture for the drain current.

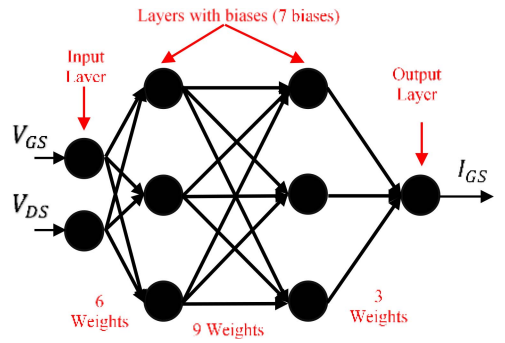

FIGURE 3. The proposed ANN architecture for the gate current.

$$k = \psi\left(\sum_{i=1}^{n} \mathcal{W}_i t_i + p \right). \quad (2)$$

Furthermore, the gate current is also modeled using 2 hidden layers based GA augmented ANN model. Each hidden layer consists of 3 neurons as shown in Fig. 3. The gate current has exponential behavior in the tried-forward region and it is mainly depends on V_{GS}. On the other hand, the drain current depends on both V_{GS} and V_{DS} and it has multiple nonlinearities in triode, pinch-off and forward region. The stronger nonlinear behavior of the drain current justify the required higher number of neurons. This also agrees with the reported analytical model in [23], which presented more complicated nonlinear formula for the drain current with respect to the exponential based formula of the gate current. The BP based algorithms are susceptible to initial values of weights and biases in order to produce fast converging and better results. To overcome this problem, GA augmented ANN models are developed. GA is well-known for its powerful exploration capability. The GA aids in the optimization of the initial weights and biases and later uses these optimal set to build the models [45]. The gate current is expressed

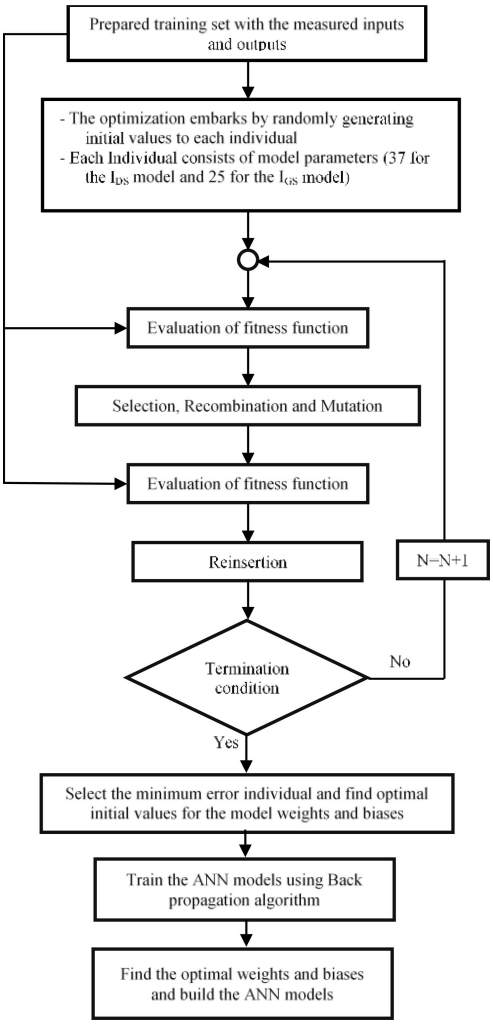

FIGURE 4. The proposed GA augmented ANN-based model.

in (3), where the terms are explained above.

$$I_{GS} = w_b^3 + \sum_{k=1}^{3} w_k tanh \times \left(w_{kb}^2 + \sum_{j=1}^{3} w_{kj} tanh\left(w_{1j} V_{GS} + w_{2j} V_{DS} + w_{bj}^1 \right) \right). \tag{3}$$

This approach is developed based on the natural selection method. Weights and biases are the models' parameter in case of ANN. Selecting the right initial values can strongly affect the overall efficiency of the ANN model, especially for large- size model with increased number of parameters. Any BP method starts the algorithm by randomly selecting initial value for the weights and biases by using appropriate probability distribution. However, if the search space is spacious, finding the right initial values to converge at global optimums is less probable. The GA can easily address this problem as it allows efficient exploration in the search space to find optimal initial values for the models' parameters. This eventually drives the algorithm to converge at the global optimums, which improves the overall accuracy of the models. Two separate models are developed. The procedure is briefed in the flow chart shown in Fig. 4. The same procedure is repeated for both models, i.e., both for modeling of I_{DS} and I_{GS}.

The only difference is the number of variables used. A brief summary of the modelling process is:

- First the training sets are prepared. The training set for the drain current is composed of input data of applied V_{GS}, V_{DS} and measured I_{DS} as an output parameter. The second model consists of the same input data but with the gate current (I_{GS}) as output parameter. The dataset is further divided into training and unseen testing sets. Almost two-third of the entire data samples are implemented for training and the remaining are used for testing. To get this distribution, for respective V_{GS}, two-third of entire V_{DS} biasing sets are included into training set and remaining into testing set. This process is applied for each model. The whole data is fed to the GA based algorithm to produce the optimal initial sets of weights and biases of the NN models.
- The optimization first generates initial population of 1000 individuals. Each individual embodies the total number of variables (37 variables: 28 weights and 9 biases for the first model and 25 variables: 18 weights and 7 biases for the second model). The variables are set within the range of -1 and 1 and are initialized randomly within this range.
- The objective function or fitness function (MSE) in (4) is defined with the objective to minimize the squared difference between the predicted values and the measured values. If K is the total number of samples then:

$$MSE = \frac{1}{K} \sum_{j=1}^{K} \left(I_{meas} - I_{pred} \right)^2. \tag{4}$$

- Once initialized properly and, regarded as parent generation, each individual is fitted using the error equation and then the fitness value is calculated. Based on the fitness value, 10 % of them are rejected (in the 1st iteration) and the remaining individuals are passed the recombination and mutation stages and eventually render the "offspring" for this second generation. These candidate solutions are evaluated by calculating their fitness errors. Then 90% of the less error are selected. The rejected 10% are replaced by the best candidate of the last generation (parents). The new formed solutions represent parents for the next third generation and the process will repeat itself. This will continue up to the predefined maximum number of iteration Nmax. In our case Nmax is set to 500 to produce the best results
- For the effective reproduction of the off-springs, a highly efficient double-edge crossover technique is exploited. The newly generated offspring's values go

TABLE 1. Gate and drain bias voltages of the drain current.

V_{GS}(V)		V_{DS}(V)	
Range	Step-size	Range	Step-size
-8 to -2 V	0.3	0 – 12 V	0.3 V
		12 – 48 V	2 V

through the same stages repeatedly, through a process called reinsertion. At the next generation, the same steps are re-evaluated until the termination condition is met.
- Then GA produces optimal set of weights and biases for the given fitness function.
- These determined sets are initial values and are used to override the BP based weights and biases. The ANN models are then trained and validated based on these initial values.

IV. PSO AUGMENTED SVR BASED MODEL

Support vector regression (SVR) is derived from Support vector machine (SVM) classification for regression-based problems. In general, for any regression-based problem the objective of the models is to minimize the squared error. However, in SVR, the objective is to minimize the absolute error term defined using certain constraints. Moreover, SVR allows the developer to define a certain error range [46]. The SVR implements linear-epsilon insensitive (e–SVM) regression, which is sometimes also referred to as $L1$-loss. It is a supervised model where the user provides a training set consisting of predictor variables and target values. The model learns the optimal weights and biases by solving the optimization problem to produce the most effective results. More details about the SVR and the implanted combined technique of SVR and PSO are provided in Appendix A.

V. GPR BASED MODEL

The Gaussian process regression (GPR) models come under the category of nonparametric kernel-based probabilistic models. A brief summary of GPR is discussed in this section. The details are available in [47]–[48]. Also more details to explain the drain and gate current modeling using GPR are provided in Appendix B.

VI. THE DEVELOPED MODEL VALIDATION

The proposed modeling approach has been applied on the DC IV measurements listed in Table 1 for a 2mm packaged GaN HEMT on Si substrate.

A. GA AUGMENTED ANN MODEL VALIDATION

The measured dataset is preprocessed before training to drive the range of features and output within $[-1, 1]$. Then the models are trained to simulate bias dependence of the DC drain and gate currents. For this purpose, the topologies depicted in Figs. 2 and 3 have been implemented and trained with the measured data. The well-known tan-sigmoid activation function is used at each layer.

TABLE 2. Mean squared error at randomly selected bias conditions (ga augmented ANN model for the drain current).

Bias	$I_{DS,DC}$ (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	1.25e-5
V_{GS} = -5.6V; V_{DS} = 8.4V	1.61e-5
V_{GS} = -1.5V; V_{DS} = 1.5V	1.35e-5
V_{GS} = -0.8V; V_{DS} = 46V	1.77e-5
V_{GS} = -0.6V; V_{DS} = 34V	1.14e-5

TABLE 3. Mean squared error at randomly selected bias conditions (GA augmented ANN model for the gate current).

Bias	I_{GS} (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	4.38e-7
V_{GS} = -5.6V; V_{DS} = 8.4V	5.38e-7
V_{GS} = -1.5V; V_{DS} = 1.5V	3.65e-7
V_{GS} = -0.8V; V_{DS} = 46V	4.31e-7
V_{GS} = -0.6V; V_{DS} = 34V	5.24e-7

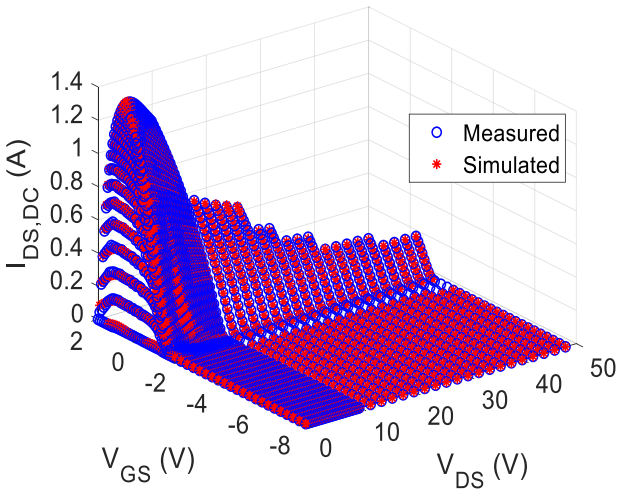

FIGURE 5. Measured and simulated drain current of 2-mm packaged GaN HEMT using GA augmented ANN model.

The GA is used to produce optimal initial values for the weights and biases. Then the models are trained using Levenburg-Marquardt (LM) BP algorithm [49] after overwriting the determined initial weights and biases. The MSE is recorded at randomly selected different biases to check the robustness of the modelled drain current and the outcome is given in Table 2. The same for the gate current is given in Table 3. Moreover, the measured and simulated plots for the drain current for a wide range of 0-48 V in Fig. 5 shows the overall behavior of the proposed model. The plots for the gate current are shown in Fig. 6. It can be observed that the GA augmented ANN provides excellent fitting for the measurements. The same results are obtained with multiple running of the program and this validates the robustness and uniqueness of the obtained solutions with respect to the conventional BP based ANN.

B. PSO AUGMENTED SVR MODEL VALIDATION

It is imperative to note that the hyper-parameters tuning plays a vital role in improving the performance of the SVR

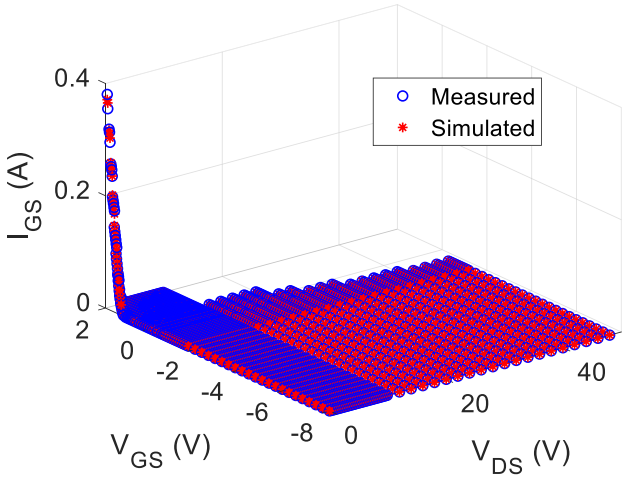

FIGURE 6. Measured and simulated gate current of 2-mm packaged GaN HEMT using GA augmented ANN model.

TABLE 4. The optimal values of hyper parameters to simulate drain current using PSO algorithm.

Kernel Function	Box Constraint (C)	Epsilon (ε)	Sigma (σ)
Gaussian	2.97	3.46e-4	1.28

TABLE 5. The optimal values of hyper parameters to simulate gate current using PSO algorithm.

Kernel Function	Box Constraint (C)	Epsilon (ε)	Sigma (σ)
Gaussian	1.23	1.99e-5	0.97

TABLE 6. Mean squared error at randomly selected bias conditions (PSO augmented SVR model for drain current.

Bias	$I_{DS,DC}$ (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	1.31e-4
V_{GS} = -5.6V; V_{DS} = 8.4V	1.83e-4
V_{GS} = -1.5V; V_{DS} = 1.5V	1.67e-4
V_{GS} = -0.8V; V_{DS} = 46V	1.24e-4
V_{GS} = -0.6V; V_{DS} = 34V	1.53e-4

based models. Keeping this in context, the PSO algorithm is used to tune the parameters. The parameters to be optimized are Box Constraint (C), width of the tube size (ε), Kernel function, standardization of data, polynomial order, and kernel scale. The Gaussian kernel is chosen after comparing the performance of linear kernel, polynomial kernel with polynomial order up to 5 in terms of MSE. The Gaussian kernel outperformed every other kernels. The standardized dataset is used. Now, the models are trained using PSO to find the optimal values of C, ε and σ. The same measurement set listed in Table 1 is used to train and validate the model. The optimal parameters of SVR models are listed in Tables 4 and 5 for the drain and gate currents, respectively. The models are again validated at randomly selected weights and biases to check the performance in different region of operations whose results are given in Tables 6 and 7. The measured and simulated plot is drawn

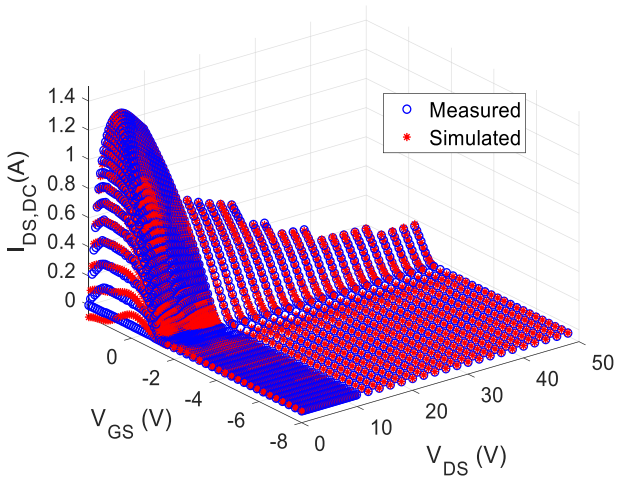

FIGURE 7. Measured and simulated drain current of 2-mm packaged GaN HEMT using PSO augmented SVR model.

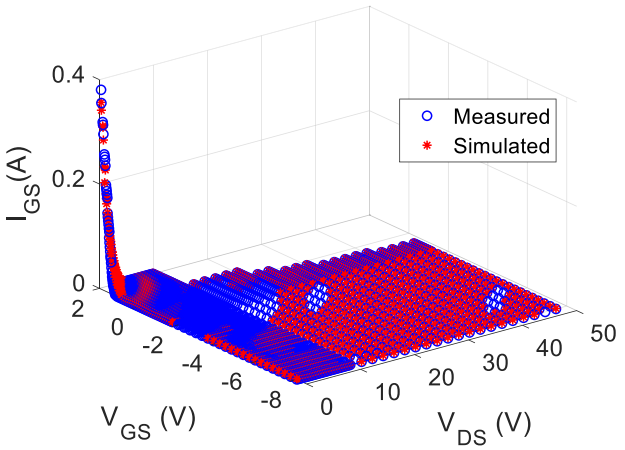

FIGURE 8. Measured and simulated gate current of 2-mm packaged GaN HEMT using PSO augmented SVR model.

TABLE 7. Mean squared error at randomly selected bias conditions (PSO augmented SVR model for gate current).

Bias	I_{GS} (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	2.28e-5
V_{GS} = -5.6V; V_{DS} = 8.4V	1.74e-5
V_{GS} = -1.5V; V_{DS} = 1.5V	1.58e-5
V_{GS} = -0.8V; V_{DS} = 46V	1.99e-5
V_{GS} = -0.6V; V_{DS} = 34V	2.89e-5

for a broad voltage range. It can be inferred from the plots shown in Figs. 7 and 8 that the models have shown good agreement. Around zero drain current the SVR has shown some deficiency. It is due to the fact that if the error lies within the ε-insensitive tube region then the SVR considers those observation to be by default as zero error and does not optimize them. This could be also observed from the higher MSE (see Tables 6 and 7) with respect to the case of GA-ANN in Tables 2 and 3.

TABLE 8. Mean squared error at randomly selected bias conditions (GPR based model for the drain current).

Bias	$I_{ds,DC}$ (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	2.14e-6
V_{GS} = -5.6V; V_{DS} = 8.4V	2.07e-6
V_{GS} = -1.5V; V_{DS} = 1.5V	1.55e-6
V_{GS} = -0.8V; V_{DS} = 46V	1.32e-6
V_{GS} = -0.6V; V_{DS} = 34V	2.77e-6

C. GPR MODEL VALIDATION

Two models, one for drain and other for gate current are developed. To reiterate, the GPR is a probability based non-parametric method. Unlike other supervised methods such as ANN, the GPR uses Bayesian optimization to create a Gaussian probability distribution over all the possible results. Generally, the flow in ANN and SVR is that we fit the data and compute the parameters using the best fitted decision boundary. In the GPR, however, the algorithm process training dataset of the current to lean their probability distribution. Then, it can predict the current at any input voltages. The datasets are standardized before training. The models then utilize the subset of data points approximation to estimate the parameters of the GPR model. This method is chosen based on the number of samples. The models also use constant basis function, exact method to predict, and Quasi-Newton method to optimize the parameters of the GPR models. These parameters are chosen by comparing MSE and number of training samples. Moreover, the key component in GPR is known as Kernel function or covariance function. The accuracy of the GPR models are greatly dependent on the kernel function as the output of the model depends on the mean and covariance matrix. Once we have the mean and covariance matrix, we can easily calculate the Gaussian distribution associated with the data. For small-sized data, the GPR has proven to work well. The squared exponential kernel is used in this work and it is expressed by (5).

$$k(t_i, t_j|\vartheta) = \sigma_f^2 exp\left(\left[\frac{-1}{2}\frac{(t_i - t_j)^T(t_i - t_j)}{\sigma_i^2}\right]\right) \quad (5)$$

where $\sigma_i$is the characteristic length scale, and σ_f is the signal standard deviation. The same training set is used to train the model listed in Table 1. To check the robustness of the models the MSE tallies are computed and listed in Tables 8 and 9 for drain and gate current, respectively. Once again, the measured and simulated plots in Figs. 9 and 10 show excellent agreement between the modelled and measured data for the entire range of voltages and thus validate the efficiency of the proposed GPR based modeling technique. As it can be noted from the values of MES in Tables 8 and 9 the GPR has comparable accuracy to the GA-ANN. However the random nature of the GPR could complicate model implementation in CAD software such as ADS.

Overall, the GA augmented ANN shows excellent agreement for both the drain and gate currents. Due to the simplicity of analytical formula of drain and gate current it has an advantage of easy implementation in ADS. Whereas the PSO based SVR is robust due to the uniqueness and its embodiment of the SRM principle. However, it has major limitation for the target values which are exactly zero as it cannot deal with the target values which are within the ε–tube region. On the other hand, the GPR has shown the best performance in terms of MSE. Unlike the SVR, it can also easily deal with zero target values. It is found that simulation time of GPR is less when compared to the PSO-SVR and GA-ANN models. Furthermore, the GPR is very robust and the algorithm needs not to run many times due to the uniqueness of the solutions at each run. This also further

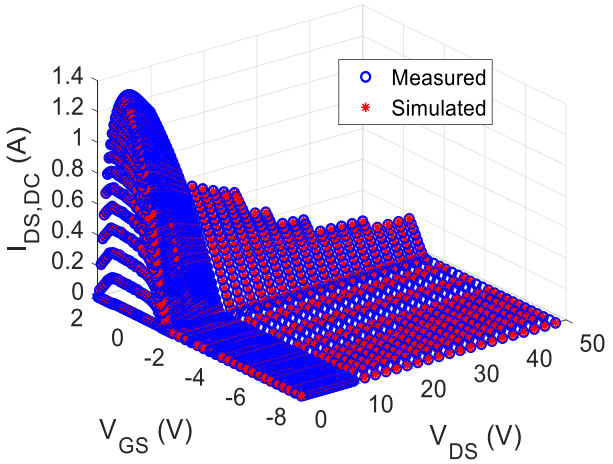

FIGURE 9. Measured and simulated drain current of 2-mm packaged GaN HEMT using GPR model.

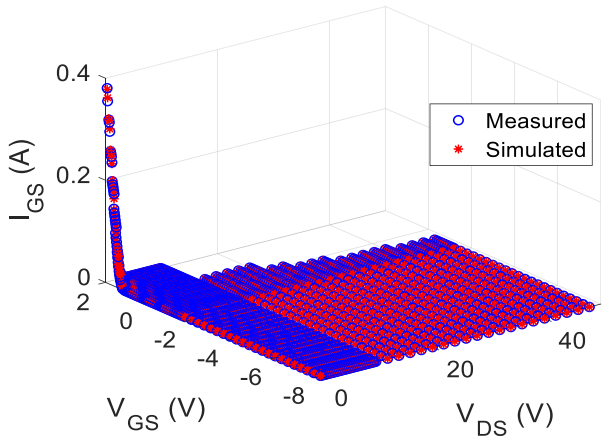

FIGURE 10. Measured and simulated gate current of 2-mm packaged GaN HEMT using GPR model.

TABLE 9. Mean squared error at randomly selected bias conditions (GPR based model for the gate current).

Bias	I_{GS} (MSE)
V_{GS} = -7.4V; V_{DS} = 3V	1.35e-7
V_{GS} = -5.6V; V_{DS} = 8.4V	1..28e-7
V_{GS} = -1.5V; V_{DS} = 1.5V	1.35e-7
V_{GS} = -0.8V; V_{DS} = 46V	1.89e-7
V_{GS} = -0.6V; V_{DS} = 34V	1.54e-7

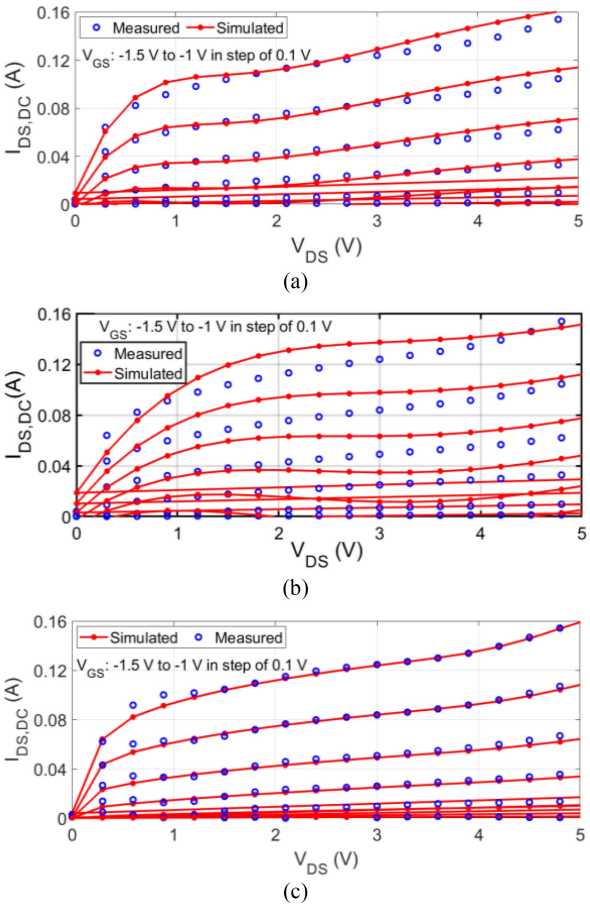

FIGURE 11. Measured and simulated drain current of 2-mm packaged GaN HEMT using: (a) GA-ANN; (b) PSO-SVR; and (c) GPR model.

TABLE 10. A compassion table of all the models.

Parameters	GA augmented ANN model	PSO augmented SVR model	GPR model
Model's simplicity	Easy to implement in ADS	Quite Difficult to implement in ADS	Difficult to implement in ADS
Computational efficiency	Computationally efficient	less computationally efficient	more computationally efficient
Average MSE	1.88x10-5 for drain and 4.47x10-7 for gate currents	1.23x10-4 for drain and 1.85x10-5 for gate currents	1.25x10-6 for drain and 1.71x10-7 for gate currents

clarifier in Fig. 11, which compares simulation of the three models for the strong nonlinear behavior in the triode region. A brief comparison table of all the proposed models is given in Table 10 and a comparison of simulation time to build the entire model is given in Table 11.

VII. LARGE-SIGNAL MODEL REALIZATION AND VALIDATION

As it was mentioned, self-heating and trapping effects induce a current dispersion that has stronger impact on the device performance and must be considered in the modeling process. The dynamic behavior of both the thermal and trapping

TABLE 11. Simulation time to build the entire models.

Model	Time required to build the model for drain current (I_{DS}) (in Seconds)	Time required to build the model for gate current (I_{GS}) (in Seconds)
GA augmented ANN model	GA simulation time: 235 ANN simulation time: 3 Total time: 238	GA simulation time: 170 ANN simulation time: 2 Total time: 172
PSO augmented SVR model	PSO simulation time: 665 SVR simulation time: 25 Total time: 690	PSO simulation time: 597 SVR simulation time: 23 Total time: 620
GPR model	44	35

represent the main source of long-term memory effect [50]. This effect results in frequency-dependency of nonlinear characteristics of the transistor amplifier and thus it should be considered by the model for efficient circuit design. To simulate these effects, a simple approach has been followed by adding two RC circuits. The thermal effect is simulated by the dynamic electro-thermal sub-circuit (see Fig. 1) with a thermal constant of 1 ms [51]. Figure 12 depicts a block diagram representation for the electro-thermal modeling of the self-heating effect, which reduces the drain current under static and quasi-static of operation; however it has negligible influence when a fast envelop varying at Radio Frequency (RF) signal stimulates the device. When using 4G signal with a modulation bandwidth equal or higher than 20 MHz, the signal' envelop is fast enough to prevent heating up the device and increasing the channel temperature; however dispersive and trapping effects due to the fast envelop can be trigged. Thus the DC drain current has to be corrected to emulate this dispersive behavior under RF modulated signals. The series $R_{rf}C_{rf}$ circuit in the drain side of the equivalent circuit in Fig. 1 is to simulate the dynamic buffer trapping, which has stronger impact with respect to the previously mentioned surface trapping. This simple approach is working well for simulating this effect under the considered bias condition [52] and it has been used here just to keep the model simplicity. An extended approach such as the presented on [25] could be used to consider the bias dependence of both surface and buffer trapping.

As can be seen in Fig. 12 the DC drain current can be modelled based on any of the proposed approaches namely the GA-ANN, PSO-SVR, and GPR. It can be deduced that the model's drain current output is multiplied by the drain voltage to calculate the DC power dissipation. This calculated power is then multiplied by frequency-dependent factor of $K_T H(\omega)$ to synthesize the self-heating induced dispersion. Here, K_T is a thermal correction factor that can be determined by comparing the DC and RF trans-conductance and output-conductance at active bias condition [17]. The term

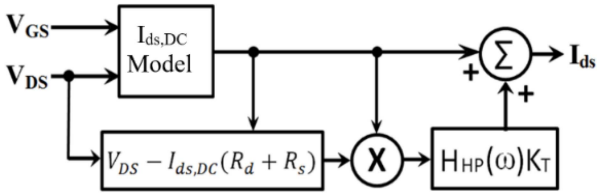

FIGURE 12. Electrothermal model of the drain current.

TABLE 12. Extrinsic elements of a 2-mm packaged GaN HEMT.

C_{gp} (pF)	C_{dp} (pF)	R_g (Ω)	R_d (Ω)	R_s (Ω)	L_g (nH)	L_d (nH)	L_s (nH)
1.39	1.65	0.77	0.91	0.38	1.58	1.3	0.14

$H(\omega)$ is a unity high-pass transfer function to account for the higher impact of self-heating due to static and quasi-static power dissipation. According to the model in Fig. 12, the drain current can be formulated as:

$$I_{ds} = I_{ds,DC}[1 + K_T H_{HP}(\omega) P_{diss}]. \quad (6)$$

As it is well know that the gate current is not frequency dependent and thus its value will be the quasi-the same under DC and RF operating conditions. Therefore, the gate current could be simulated using one of the GA-ANN, PSO-SVR or GPR model based on DC measured data. As it was presented in Figs. 6, 8 and 10 all models showed very good fitting to the measurements. As it was mentioned, the extrinsic elements of the equivalent circuit model of Fig. 1 have been extracted based on cold S-parameter measurements and using the same procedure reported in [17]. Table 12 lists the extracted elements of the investigated 2-mm device. These extrinsic elements are then de-embedded from active (hot) S-parameter measurements to model the intrinsic transistor and deduce the values of its elements. The intrinsic elements C_{gs} and C_{gd} are then extracted by means of curve fitting of the intrinsic Y-parameters [17]. The bias dependence of C_{gs} and C_{gd} has been simulated using simple single hidden layer ANN models of three neurons. Fig. 12 shows the extracted values with the models fitting. Under bias conditions of the typical high efficient amplifier, C_{ds} is almost constant. For that reason and to keep the simplicity the model, C_{ds} is considered as a bias-independent element. The equivalent circuit model of Fig. 1 with its extracted parameters is implemented in the Advanced Design System (ADS) software. The GA-ANN models of $I_{DS,DC}$ and I_{GS} are implemented analytically using the closed-form formulations given in (1) and (3), respectively. The implemented model is validated by single-tone large-signal RF characterization under different bias conditions.

As it can be seen in Fig. 14, the simulation results agree well with the measurements. The developed model also shows a greater rate of convergence and a shorter simulation time with respect to the table-based model for the same device [17]. This of course should be expected because of the discrete nature of the table-based model as compared to the proposed continuous model.

The developed model has been also used to simulate an inverse class-F switching-mode power amplifier (see

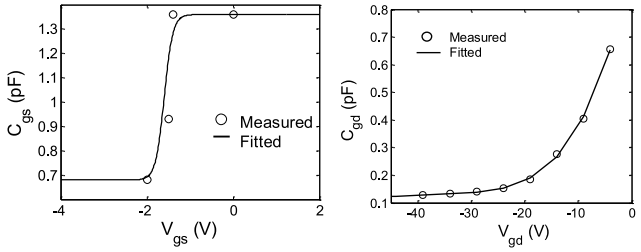

FIGURE 13. Measured and Simulated intrinsic capacitances at $V_{GS} = -4$ V.

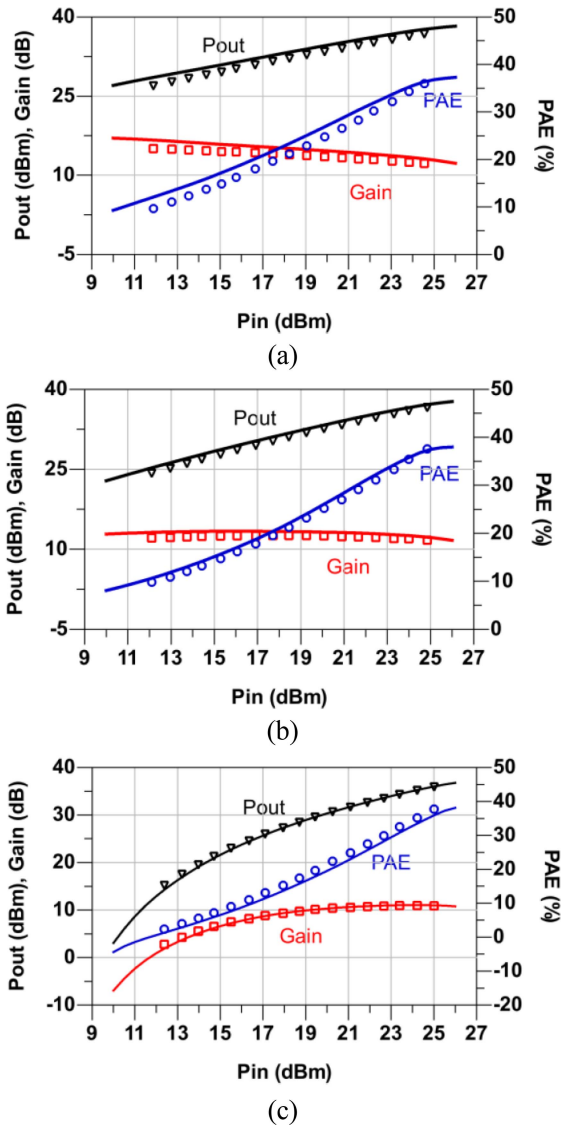

FIGURE 14. Simulated (lines) and measured (symbols) power sweep for 50-O terminated 2-mm packaged GaN HEMT at 2.35 GHz and under the bias conditions of: (a) $V_{GS} = -1$ V and $V_{DS} = 28$ V, (b) $V_{GS} = -1.6$ V and $V_{DS} = 28$ V and (c) $V_{GS} = -2.5$ V and $V_{DS} = 28$ V.

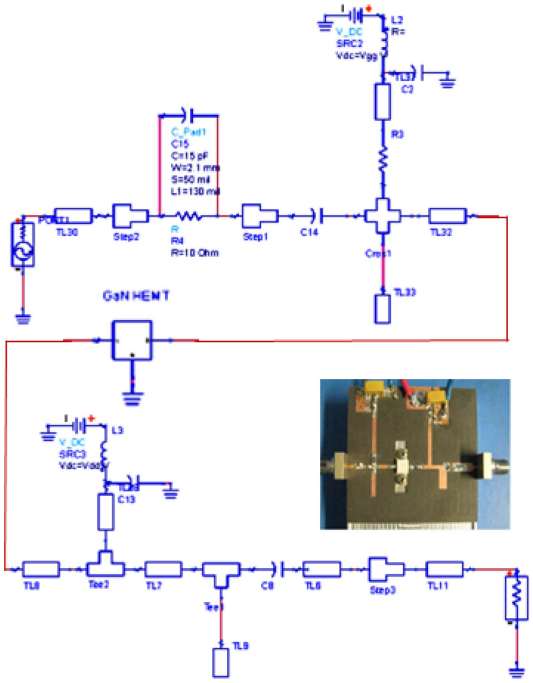

FIGURE 15. ADS simulated and fabricated inverse class-F power amplifier.

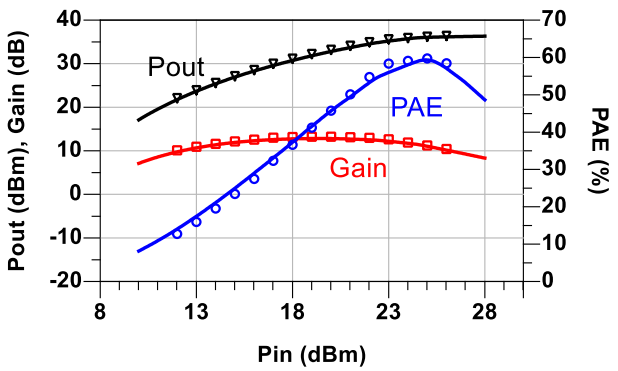

FIGURE 16. Simulated (lines) and measured (symbols) power sweep for 50-O terminated invers class-F power amplifier at 2.35 GHz and under the bias conditions of $V_{GS} = -2.5$ V and $V_{DS} = 28$ V.

Fig. 15) based on the same considered device. This amplifier has been previously designed using a Table-Based (TB) model for the same device and the results were reported in [9]. The amplifier was adopted to operate at 2.35 GHz frequency. It was also well matched with 50-O source and load conditions. The amplifier is biased below pinch-off ate $V_{GS} = -2.5$ V and $V_{DS} = 28$ V. The amplifier was simulated in ADS using a Computer with 3.4 GHz core i7 CPU, 16 GB RAM and 64-bit operating system. Fig. 16 shows simulated and measured output power, power-added-efficiency, and gain of the amplifier. As it can be seen, the model can accurately simulate the amplifier and efficiently predict the peak PAE. The model also showed faster simulation with higher rate of convergence with respect to the TB model in [9]. The total simulation time of the amplifier using the TB model is 27.5 s, while it is 9.6 s using the proposed model. The simulation time is reduced by nearly three times, which is very useful especially for simulating more complicated circuits such as multi-stage power amplifiers.

VIII. CONCLUSION

In this article, efficient large-signal modeling techniques using machine learning based approaches are developed and implemented for GaN HEMT. The proposed approaches employed GA augmented ANN, PSO augmented SVR, and GPR based models to develop analytical electro-thermal models for the drain and gate currents of the GaN HEMT. The developed models are compared in terms of MSE and simulation time. All the models are validated using simulation plots and excellent agreement is obtained. It has been observed that the PSO-SVR training algorithm is negatively impacted by the drawback of not accounting zero target values. The GA augmented ANN model, owing to its simplicity, was then utilized for investigation and validation under single-tone large-signal RF excitations. In summary, the GA augmented ANN and GPR based models offers simple analytical formulation and can be therefore readily adapted in commercial CAD tools. It has been found that a reduction of around three times in simulation time with higher rate of convergence could be obtained using the proposed modeling with respect to the table-based technique. This validates the applicability of the proposed modeling approaches for simulating more complicated nonlinear circuits.

APPENDIX A

As it was mentioned, the SVR model is based on using box constraint or regularization parameters, which works as a trade-off between making the weights ($\|\omega^2\|$) small (to make the margin large) and ensuring that each example has functional margin of at least 1. In other words, it is a penalty factor, which penalizes the observations going beyond the epsilon margin range. The e-insensitive loss function does not notice and overlook the error values, which are within the range of e by considering them as zero. Another prominent attribute of SVR is the kernel trick, which in principle, converts the non-linear low dimension- dataset to linear high-dimensional dataset and computes dot products to produce a scalar output and helps in dealing with complex problems. Assume t_n is a multivariate set of N observations and k_n is the corresponding response value. The objective of SVR model is to find $f(x)$, which does not deviate more than e from t_n for each training point x. The SVR formulation for a non-linear model is given in (7). Here, ω is the weight and p is the bias, and $\phi(x)$ is the kernel trick or kernel matrix.

$$f(x) = \langle \omega, \phi(x) \rangle + p. \tag{7}$$

To insure that $f(x)$ is as flat as possible, find $f(x)$ with minimum norm value ($\omega^T \omega$). It can be converted as a convex optimization problem to minimize, which can be solved using Lagrange multiplier optimization technique, and to find the optimal hyperplane and to convert a non-convex problem

to a convex solvable optimization problem. Generally, we define functional and geometric margins. By using the scaling constraints on weights and biases and altering the maximizing problem to a minimization problem we reach (8).

$$Minimize\left(\frac{1}{2}\left|\left|\omega^2\right|\right|\right)$$

with the constraints:

$$\left\|k_n-\left(t_n'\omega+p\right)\right\|\leq\varepsilon. \tag{8}$$

There is a possibility that no such term exists which follows the constraint defined by (8). To deal with such a situation, other terms known as slack variables (ξ_i, ξ_i^*) are added to the equation. The expression, including the slack variables, known as primal formula for SVR is given in (9) with subject to conditions in (10). The primal formula can be solved by converting it to a Lagrange dual problem [46].

$$Minimize\left(\frac{1}{2}\left|\left|\omega^2\right|\right|+C\sum_{i=1}^{n}\left(\xi_i+\xi_i^*\right)\right) \tag{9}$$

$$k_n-(\langle\omega,t_n\ \rangle+p)\leq\varepsilon+\xi_i$$
$$(\langle\omega,t_n\rangle+p)-k_n\leq\varepsilon+\xi_i^*$$
$$\xi_i\geq 0\ and\ \xi_i^*\geq 0. \tag{10}$$

Then the predictive function ($f(x)$) for the nonlinear primal function can be completely describe in terms of support vectors.

$$f(x)=\sum_{n=1}^{N}\left(\alpha_n^*-\alpha_n\right)G(t_n,x)+p \tag{11}$$

The terms α_n^* and α_n in brackets are weight coefficients of support vectors, p is the bias and $G(\cdot)$ is the gram matrix. Then the formulations for drain current can be expressed by (12). Here, t_n is the support vectors, N is the number of support vectors, σ is the Gaussian kernel parameter and p is the bias. Similarly, equation for the gate current can be formulated by (13).

$$I_{DS,DC}=\sum_{n=1}^{N}\left(\alpha_n^*-\alpha_n\right)exp \times\left(\frac{(t_n-\{[V_{DS},V_{GS}]\})^T(t_n-\{[V_{DS},V_{GS}]\})}{2\sigma^2}\right)+p \tag{12}$$

$$I_{GS}=\sum_{n=1}^{N}\left(\alpha_n^*-\alpha_n\right)exp \times\left(\frac{(t_n-\{[V_{DS},V_{GS}]\})^T(t_n-\{[V_{DS},V_{GS}]\})}{2\sigma^2}\right)+p \tag{13}$$

One of the main reasons of degradations of the SVR's predictive ability is the inappropriate selection of the parameters. Different parameters affect the performance in different ways and hence it is important to tune and optimize the central parameters to improve the predictive ability of the model. This article makes use of PSO to tune the parameters. The PSO comes under the category of global optimization techniques and thus is not susceptible to local solutions but more inclined towards the global solutions of a given optimization problem. It is fast and computationally less intensive as opposed to complex methods such as GA and therefore works excellently when optimization of fewer parameters is envisaged. As discussed earlier, the box constraint (C) is an essential parameter, which directly supervises the overfitting and under fitting of the model. For a large value of C it will lead the model to overfit and for very small value of C it will lead the model to underfit, so finding a middle value is very crucial. The ε-insensitive loss function, defines the margin of tolerance. This directly influences the number of support vectors. As per (11), the output of the SVR model is mainly dependent on the quantity and quality of the support vectors. So, optimizing the error value is of utmost importance. Moreover, selecting the right kernel for the problem greatly influences the overall accuracy. For SVR, there are many kernels available such as Gaussian kernel, linear kernel, and polynomial kernel. It is also possible to design a custom kernel function to deal with the specific problems. However, Gaussian kernel is used in the current work. Here, two models are developed, one to simulate drain current and the second to depict the gate current. Both models utilize the principle of SVR and make use of PSO to optimize the models' parameters. The box constraint (C) and kernel scale or sigma are initialized randomly and set to traverse within the range of $[10^{-3}, 10^3]$ for both the models. The ε-insensitive loss function, is also chosen randomly and allowed to take values between $[10^{-5}, 10^5]$ for the drain current and $[10^{-6}, 10^6]$ for the gate current. This process is repeated many times until the optimal values of these terms are obtained. In fact, the PSO is a model developed by understanding the social behavior of birds flocking or fish schooling [37], where each particle possesses its own position and velocity. All the particles share a common objective either to seek food or avoid predators or to find the adaptable environmental parameters. In order to accomplish the objective each particle adjust their own positions and velocities which can be mathematically explained. Moreover, they also keep track of the best position explored by them. These local and best position together decide the velocity of each particle.

The proposed methods are briefly described in the flow chart of Fig. 17, and can be summarized as follows:

- The training of PSO starts by feeding the measured data containing inputs and output to the algorithm. Upon receiving the data, the algorithm creates initial population of particles (100 particles). Each particle contains 3 variables for the parameters C, ε and σ.
  These variables are assigned random values within the above-mentioned ranges.

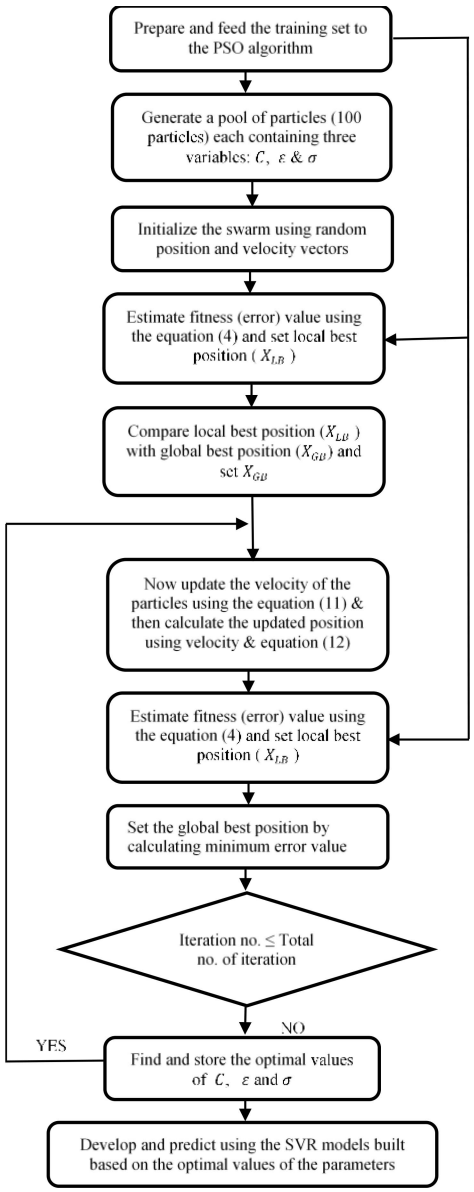

FIGURE 17. The proposed PSO augmented SVR-based model.

- The next step is the evaluation of fitness function. Its value is calculated using (4). Based on the fitness value, the local best position of individual particle is set. Moreover, by comparing the positions of the particles, the global best position is determined. Local best position and global best position continue to be updated after each successful iteration until the termination condition is met.
- Each particle updates its velocity based on individual best position and global best position. The velocity of each particle is updated using (14) and (15).

$$v_{n+1} = w*v_n + c_1*r_1*(X_{LB} - x_n) + c_2*r_2 * (X_{GB} - x_n) \quad (14)$$

$$x_{n+1} = x_n + v_{n+1} \quad (15)$$

where, r_1 and r_2 between 0 and 1 preferably obtained using stochastic methods. v_{n+1} is the updated velocity and v_n is the previous or memory velocity. Similarly, x_{n+1} is the updated position and x_n is the previous or memory position. X_{LB} and X_{GB} are the local best and global best positions, respectively. The terms w, c_1 and c_2 are the inertia weight factor, self-confidence factor, and swarm confidence factor. The inertia factor is calculated using (16).

$$w = w_{max} - \left(\frac{(w_{max} - w_{min})}{Max_{Iteration}} * Iteration \right). \quad (16)$$

Here, w_{max} and w_{min} are the maximum and minimum range for the inertia factor values. These values can be greatly altered by the internal features associated with the processes used to update the velocity.

- The updated velocity is then used to update the new local best and global best positions. This process is repeated for each particle, at each generation, until the process stops by the termination condition set by the user. In our case, the algorithm runs for 250 iterations.
- After the termination condition is reached, the algorithm provides the optimal set of the three above listed parameters. These values are then used to build the SVR model. The model is trained and validated in a later section.

APPENDIX B

As it was mentioned, GPR models come under the category of nonparametric kernel-based probabilistic models. Assume the training set, $\mathcal{D} = \{t_j, y_j\}, j = 1, 2, 3, \ldots, m$, where the feature matrix $t_j \in \mathbb{R}^d$, the target matrix $y_j \in \mathbb{R}$, from an unknown distribution. A GPR model assumes a Gaussian process a priori, which can be characterized in terms of mean, $m(t)$ and covariance function, $k(t, t')$ as follows:

$$f(t) \sim GP(m(t),\ k(t, t')). \quad (17)$$

Furthermore, a linear regression model of GPR problem can be formulated by incorporating the noise or error information as expressed in (18). Here, $\varepsilon \sim N(0, \sigma^2)$ that is ε is independently and identically distributed (*i.i.d.*) Gaussian noise with mean zero and variance σ^2. The $f(t)$ is assumed to be Gaussian distribution and hence the y, the observed output, can form a Gaussian process expressed by (19). The term δ_{ij} is Kronecker delta function, when $i = j, \delta_{ij} = 1$. Subsequently, the covariance function is expressed in (20). In this expression, $C(T, T)$ denotes covariance matrix of NxN, I denotes the unit matrix of NxN, $K(T, T)$ known as Gram matrix which denotes nuclear matrix NxN, that contains K_{ij}=$k(t^i t^j)$ as its elements. It can be safely stated from the Gaussian process that the collection of training points and test points are joint multivariate Gaussian distributed and their distribution can be expressed by (21). Therefore, primary GPR equations are in (22). Here, $\overline{f_*}$ in (23) is the predicted mean output of the GPR model on output vector.

$$y = f(t) + \varepsilon \quad (18)$$

$$y = GP\Big(m(t),\ k\big(t,t'\big) + \sigma_m^2\delta_{ij}\Big) \tag{19}$$

$$C(T,T) = K(T,T) + \sigma_m^2 I \tag{20}$$

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix},\ \begin{bmatrix} K(T,T)+\sigma_m^2 I & K(T,T_*) \\ K(T_*,T) & K(T_*,T_*) \end{bmatrix}\right) \tag{21}$$

$$f_*\big|T,y,T_* \sim N\big(\overline{f_*},\ \overline{\Sigma_*}\big) \tag{22}$$

$$\overline{f_*} = \mu_* + K(T_*,T)\Big[K(T,T)+\sigma_m^2 I\Big]^{-1}(y-\mu) \tag{23}$$

$$\Sigma_* = K(T_*,T_*) - K(T_*,T)\Big[K(T,T)+\sigma_m^2 I\Big]^{-1}K(T,T_*). \tag{24}$$

For the gate and drain current, T is a multivariate set of inputs composed of V_{DS} and V_{GS} given by (25).

$$T = \left\{\left[V_{DS}^i, V_{GS}^i\right]\right\};\ \ i = 1,2,\ldots,n. \tag{25}$$

where V_{Ds}^i and V_{GS}^i are column vectors where each row defines a new observation and n is the total number of observations in the test set. To sample a function, it first computes the covariance between all observations of V_{DS} and V_{GS}, denoted by $K(T,T)$, and usual prior mean $m(T)$ as zero. Afterwards, it defines the sample function (f_*) as:

$$f_* = N(0, K(T,T)) \tag{26}$$

Suppose the training set, $\mathcal{D}_t = \{X_t, y_t\}$; where X_t is the multivariate set of inputs composed of training observations of gate and drain bias voltages and y_t is the drain or gate current. By using the standard results, the conditional distribution $p(f_*|\mathcal{D}_t, T)$ is computed using (27) and (28).

$$m(T) = K(T,X_t)\Big[K(X_t,X_t)+\sigma_m^2 I\Big]^{-1} y_t \tag{27}$$

$$k_t\big(T,T'\big) = k\big(T,T'\big) - k(T,X_t)\Big[K(X_t,X_t)+\sigma_m^2 I\Big]^{-1} k\big(X_t,T'\big). \tag{28}$$

To predict f_*, it uses simply mean function given in (27) or sample function from the GP with this mean function and kernel (28) as described before. Now, the predicted $I_{Ds,DC}$ and I_{GS} can be analytically calculated using (29) and (30) using the mean and kernel defined in (24) and (25).

$$I_{Ds,DC} = N\big(m(T),\ k_t\big(T,T'\big)\big) \tag{29}$$

$$I_{GS} = N\big(m(T),\ k_t\big(T,T'\big)\big). \tag{30}$$

REFERENCES

[1] M. S. Cho, J. H. Seo, S. H. Lee, H. S. Jang, and I. M. Kang, "Fabrication of AlGaN/GaN fin-type HEMT using a novel T-gate process for improved radio-frequency performance," *IEEE Access*, vol. 8, pp. 139156–139160, 2020.

[2] M. Noweir, M. Helaoui, W. Tittel, and F. M. Ghannouchi, "Carrier aggregated radio-over-fiber downlink for achieving 2Gbps for 5G applications," *IEEE Access*, vol. 7, pp. 3136–3142, 2018.

[3] F. Thome, A. Leuther, M. Schlechtweg, and O. Ambacher, "Broadband high-power W-band amplifier MMICs based on stacked-HEMT unit cells," *IEEE Trans. Microw. Theory Techn.*, vol. 66, no. 3, pp. 1312–1318, Mar. 2018.

[4] K. Rawat, M. S. Hashmi, and F. M. Ghannouchi, "Double the band and optimize," *IEEE Microwave Mag.*, vol. 13, no. 2, pp. 69–82, Mar./Apr. 2012.

[5] K. Rawat, M. S. Hashmi, and F. M. Ghannouchi, "Dual-band RF circuits and components for multi-standard software defined radios," *IEEE Circuits Syst. Mag.*, vol. 12, no. 1, pp. 12–32, 1st Quart., 2012.

[6] U. K. Mishra, P. Parikh, and Y.-F. Wu, "AlGaN/GaN HEMTs-an overview of device operation and applications," *Proc. IEEE*, vol. 90, no. 6, pp. 1022–1031, Jun. 2002.

[7] N. Keshmiri, D. Wang, B. Agrawal, R. Hou, and A. Emadi, "Current status and future trends of GaN HEMTs in electrified transportation," *IEEE Access*, vol. 8, pp. 70553–70571, 2020.

[8] F. Zeng *et al.*, "A comprehensive review of recent progress on GaN high electron mobility transistors: Devices, fabrication and reliability," *Electronics*, vol. 7, no. 12, p. 377, 2018.

[9] A. Jarndal, L. Arivazhagan, and D. Nirmal, "On the performance of GaN-on-silicon, silicon-carbide, and diamond substrates," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 6, 2020, Art. no. e22196.

[10] A. Khusro, M. S. Hashmi, A. Q. Ansari, A. Mishra, and M. Tarique, "An accurate and simplified small signal parameter extraction method for GaN HEMT," *Int. J. Circuit Theory Appl.*, vol. 37, no. 6, pp. 941–953, Jun. 2019.

[11] R. S. Pengelly, S. M. Wood, J. W. Milligan, S. T. Sheppard, and W. L. Pribble, "A review of GaN on SiC high electron-mobility power transistors and MMICs," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 6, pp. 1764–1783, Jun. 2012.

[12] S. Colangeli, A. Bentini, W. Ciccognani, E. Limiti, and A. Nanni, "GaN-based robust low-noise amplifiers," *IEEE Trans. Electron Devices*, vol. 60, no. 10, pp. 3238–3248, Oct. 2013.

[13] K. J. Chen *et al.*, "GaN-on-Si power technology: Devices and applications," *IEEE Trans. Electron Devices*, vol. 64, no. 3, pp. 779–795, Mar. 2017.

[14] F. M. Ghannouchi and O. Hammi, "Behavioral modeling and predistortion," *IEEE Microw. Mag.*, vol. 10, no. 7, pp. 52–64, Dec. 2009.

[15] F. M. Ghannouchi and M. S. Hashmi, *Load-Pull Techniques With Applications to Power Amplifier Design*. Dordrecht, The Netherlands: Springer, 2012.

[16] X. Cheng, M. Li, and Y. Wang, "Physics-based compact model for AlGaN/GaN MODFETs with close-formed I–V and C–V characteristics," *IEEE Trans. Electron Devices*, vol. 56, no. 12, pp. 2881–2887, Dec. 2009.

[17] A. Jarndal *et al.*, "Large-signal model for AlGaN/GaN HEMTs suitable for RF switching-mode power amplifiers design," *Solid-State Electron.*, vol. 54, no. 7, pp. 696–700, Jul. 2010.

[18] A. Jarndal, A. Z. Markos, and G. Kompa, "Improved modeling of GaN HEMTs on Si substrate for design of RF power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 59, no.3, pp. 644–651, Mar. 2011.

[19] J. B. King and T. J. Brazil, "Nonlinear electrothermal GaN HEMT model applied to high-efficiency power amplifier design," *IEEE Trans. Microw. Theory Techn.*, vol. 61, no. 1, pp. 444–454, Jan. 2013.

[20] S. Ghosh, A. Dasgupta, S. Khandelwal, S. Agnihotri, and Y. S. Chauhan, "Surface-potential-based compact modeling of gate current in AlGaN/GaN HEMTs," *IEEE Trans. Electron Devices*, vol. 62, no. 2, pp. 443–448, Feb. 2015.

[21] S. Khandelwal *et al.*, "robust surface-potential-based compact model for GaN HEMT IC design," *IEEE Trans. Electron Devices*, vol. 60, no. 10, pp. 3216–3222, Oct. 2013.

[22] A.-D. Huang, Z. Zhong, W. Wu, and Y.-X. Guo, "An artificial neural network-based electrothermal model for GaN HEMTs with dynamic trapping effects consideration," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 8, pp. 2519–2528, Aug. 2016.

[23] A. Jarndal and F. M. Ghannouchi, "Improved modeling of GaN HEMTs for predicting thermal and trapping-induced-kink effects," *J. Solid-State Electron.*, vol. 123, pp. 19–25, Sep. 2016.

[24] A. Khusro, S. Husain, M. S. Hashmi, and A. Q. Ansari, "Small signal behavioral modeling technique of GaN high electron mobility transistor using artificial neural network: An accurate, fast, and reliable approach," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 4, 2020, Art. no. e22112.

[25] A. Jarndal, "On neural networks based electrothermal modeling of GaN devices," *IEEE Access*, vol. 7, pp. 94205–94214, 2019.

[26] A. H. Jarndal and S. Muhaureq, "A particle swarm neural networks electrothermal modeling approach applied to GaN HEMTs," *J. Comput. Electron.*, vol. 18, pp. 1272–1279, Aug. 2019. [Online]. Available: https://doi.org/10.1007/s10825-019-01397-1
[27] X. Du *et al.*, "ANN-based large-signal model of AlGaN/GaN HEMTs with accurate buffer-related trapping effects characterization," *IEEE Trans. Microw. Theory Techn.*, vol. 68, no. 7, pp. 3090–3099, Jul. 2020.
[28] A. Khusro, M. S. Hashmi, A. Q. Ansari, and M. Auyenur, "A new and reliable decision tree based small-signal behavioral modeling of GaN HEMT," in *Proc. IEEE Midwest Symp. Circuits Syst. (MWSCAS)*, Dallas, TX, USA, Aug. 2019, pp. 303–306.
[29] G. Orengo, P. Colantonio, F. Giannini, M. Pirola, V. Camarchia, and S. D. Guerrieri, "Advanced neural network techniques for GaN-HEMT dynamic behavior characterization," in *Proc. IEEE 36th Eur. Microw. Integr. Circuits Conf.*, Manchester, U.K., Sep. 2006, pp. 10–13.
[30] J. Xu, D. Gunyan, M. Iwamoto, A. Cognata, and D. E. Root, "Measurement-based non-quasi-static large-signal FET model using artificial neural networks," in *IEEE MTT-S Int. Microw. Symp. Dig.*, San Francisco, CA, USA, Jun. 2006, pp. 469–472.
[31] Q.-J. Change and K. C. Gupta, *Neural Networks for RF and Microwave Design*. Norwood, MA, USA: Artech House, 2000.
[32] M. Gori and A. Tesi, "On the problem of local minima in back-propagation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 1, pp. 76–86, Jan. 1992.
[33] X. G. Wang, Z. Tang, H. Tamura, M. Ishii, and W. D. Sun, "An improved backpropagation algorithm to avoid the local minima problem," *Neurocomputing*, vol. 56, pp. 455–460, Jan. 2004.
[34] Q. Dai and N. Liu, "Alleviating the problem of local minima in Backpropagation through competitive learning," *Neurocomputing*, vol. 94, pp. 152–158, Oct. 2012.
[35] M. Locatelli and F. Schoen, *Global optimization: Theory, Algorithms and Applications*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2013.
[36] A. S. Hussein and A. H. Jarndal, "Reliable particle-swarm-optimization based parameter extraction method applied to GaN HEMTs," in *Proc. 16th Mediterr. Microw. Symp.*, Abu Dhabi, United Arab Emirates, 2016, pp. 1–4.
[37] A. S. Hussein and A. H. Jarndal, "Reliable hybrid small-signal modeling of GaN HEMTs based on particle-swarm-optimization," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 37, no. 9, pp. 1816–1824, Sep. 2018.
[38] W. Zhao, T. Tao, E. Zio, and W. Wang, "A novel hybrid method of parameters tuning in support vector regression for reliability prediction: Particle swarm optimization combined with analytical selection," *IEEE Trans. Rel.*, vol. 65, no. 3, pp. 1393–1405, Sep. 2016.
[39] A. Dai, X. Zhou, H. Dang, M. Sun, and Z. Wu, "Intelligent modeling method for a combined radiation-convection grain dryer: A support vector regression algorithm based on an improved particle swarm optimization algorithm," *IEEE Access*, vol. 6, pp. 14285–14297, 2018.
[40] J. Cai, J. King, C. Yu, J. Liu, and L. Sun, "Support vector regression based behavioral modeling technique for RF power transistors," *IEEE Microw. Compon. Lett.*, vol. 28, no. 5, pp. 428–430, May 2018.
[41] A. Khusro, M. S. Hashmi, and A. Q. Ansari, "Enabling the development of accurate intrinsic parameter extraction model for GaN HEMT using support vector regression (SVR)," *IET Microw. Antennas Propag.*, vol. 13, no. 9, pp. 1457–1466, Jul. 2019.
[42] A. Bilal, S. M. Hamza, Z. Taj, and S. Salamat, "Multi-frequency analysis of Gaussian process modelling for aperiodic RCS responses of a parameterised aircraft model," *IET Radar Sonar Navig.*, vol. 14, no. 7, pp. 1061–1067, Jul. 2020.
[43] Q. H. Zhang and Y.-Q. Ni, "Improved most likely heteroscedastic Gaussian process regression via Bayesian residual moment estimator," *IEEE Trans. Signal Process.*, vol. 68, pp. 3450–3460, May 2020.
[44] C. Huang, Z. Zhao, L. Wang, Z. Zhang, and X. Luo, "Point and interval forecasting of solar irradiance with an active Gaussian process," *IET Renewable Power Gener.*, vol. 14, no. 6, pp. 1020–1030, Apr. 2020.
[45] M. Melanie, *An Introduction to Genetic Algorithms*. Cambridge, MA, USA: MIT Press, 1996.
[46] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Nov. 2003.
[47] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.
[48] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Nov. 2011.
[49] J. Shawash and D. R. Selviah, "Real-time nonlinear parameter estimation using the Levenberg–Marquardt algorithm on field programmable gate arrays," *IEEE Trans. Ind. Electron.*, vol. 60, no. 1, pp. 170–176, Jan. 2013.
[50] F. M. Barradas, L. C. Nunes, T. R. Cunha, P. M. Lavrador, P. M. Cabral, and J. C. Pedro, "Compensation of long-term memory effects on GaN HEMT-based power amplifiers," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 9, pp. 3379–3388, Sep. 2017.
[51] E. Kohn, I. Daumiller, M. Kunze, J. Van Nostrand, J. Sewell, and T. Jenkins, "Switching behaviour of GaN-based HFETs: Thermal and electronic transients," *Electron. Lett.*, vol. 38, no. 12, pp. 603–605, Jun. 2002.
[52] C. Camacho-Peñasola and C. S. Aitchison, "Modelling frequency dependence of output impedance of a microwave MESFET at low frequencies," *Electron. Lett.*, vol. 21, no. 12, pp. 528–529, Jun. 1985.