Postprint of: Kozieł S., Calik N., Mahouti P., Belen M., Accurate Modeling of Antenna Structures by Means of Domain Confinement and Pyramidal Deep Neural Networks, IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION (2021), pp. 1-15, DOI: 10.1109/TAP.2021.3111299

Accurate Modeling of Antenna Structures by Means of Domain Confinement and Pyramidal Deep Neural Networks

Slawomir Koziel, *Senior Member*, *IEEE*, Nurullah Çalık, Peyman Mahouti, and Mehmet A. Belen

***Abstract*—The importance of surrogate modeling techniques has been gradually increasing in the design of antenna structures over the recent years. Perhaps the most important reason is a high cost of full-wave electromagnetic (EM) analysis of antenna systems. Although imperative in ensuring evaluation reliability, it entails considerable computational expenses. These are especially pronounced when carrying out EM-driven design tasks such as geometry parameter tuning or uncertainty quantification, both requiring repetitive simulations. Conducting some of the design procedures, e.g., global search or yield optimization, directly at the level of simulation models may be prohibitive. The employment of fast replacement models (or surrogates) may alleviate these difficulties; yet, accurate modeling of antenna structures faces its own challenges. The two major obstacles are the curse of dimensionality, manifesting itself in a rapid growth of the number of training data samples necessary to render a reliable model (as a function of the number of antenna parameters), and high nonlinearity of antenna characteristics. Recently, the concept of performance-driven modeling has been introduced, where the modeling process is focused on a small region of the parameters space, which contains high-quality designs with respect of the considered performance figures. The most advanced variation of this class of methods is nested kriging, where both the model domain and the surrogate itself are constructed through kriging interpolation. Domain confinement is realized using a set of pre-optimized reference designs, and allows for significant improvement of the model predictive power while using a limited number of training data samples. In this work, the constrained modeling concept is coupled with a novel pyramidal deep regression network (PDRN) surrogate, which offers improved handling of highly-nonlinear antenna responses. Three examples of microstrip antennas are used to demonstrate the advantages of constrained PDRN metamodels over the nested kriging surrogates with the (average) accuracy improved by a factor of two without increasing the training data set cardinality.**

***Index Terms*—Surrogate modeling, antenna design, domain confinement, nested kriging, deep neural networks.**

I. Introduction

Design of antenna systems is a complex and multifaceted task, which requires meticulous development of the structure geometry [1], [2], but also careful tuning of its parameters [3]. Depending on the application area and performance demands imposed upon the device, the antenna topology may incorporate various modifications of the base geometries (e.g., stubs of various shapes [4], ground plane [5] and radiator slots [6], defected ground structures [7], shorting pins [8], etc.), require appropriate allocation of its components (e.g., to excite orthogonal modes in circularly polarized antenna [9], improve radiator isolation in MIMO systems [10]), or a construction of compound devices, e.g., frequency selective surfaces allocated over a radiator for directivity improvement [11], etc. The aforementioned and other design techniques lead to increasingly complex designs described by large numbers of parameters, but also phenomena that affect the system operation (mutual coupling [12], the presence of connectors [13], radomes [14]). Accounting for these in an adequate manner can only be done through full-wave electromagnetic (EM) analysis, which has become ubiquitous in the design of modern antenna structures [15]-[19]. Evaluation accuracy and versatility perhaps the most appealing features of EM simulation tools, whereas potentially long simulation time is a disadvantage. The latter becomes problematic whenever repetitive analyzes are necessary. This is a common issue for most EM-driven design task, including parameter tuning [20], multi-objective design [21], statistical analysis [22], or tolerance-aware optimization [23].

The literature offers various tools for alleviating the difficulties incurred by the aforementioned high-cost issue. Some of these are strictly algorithmic approaches, both intrusive (e.g., adjoint sensitivities [24]), and non-intrusive (e.g., gradient-based procedures accelerated by sparse sensitivity updates [25], [26]); others exploit customized EM solvers tailored to particular classes of antenna structures [27], [28]. Notwithstanding, a large part of recent developments rely on surrogate modeling techniques [29]-[35]. By far, the most popular ones are approximation models, which are fast to evaluate and easily accessible [36], [37]. Because of being data-driven, these models are versatile and readily transferrable between various application domains. Widely used techniques include polynomial regression [38], kriging [39], radial-basis functions [40], neural networks [41], [42], support vector regression [43], Gaussian process regression [44], or polynomial chaos expansion [45].

The manuscript was submitted on March 31, 2021. This work was supported in part by the Icelandic Centre for Research (RANNIS) Grant 217771, and by National Science Centre of Poland Grant 2018/31/B/ST7/02369.

S. Koziel is with Engineering Optimization and Modeling Center of Reykjavik University, Reykjavik, Iceland (e-mail: koziel@ru.is). S. Koziel is also with Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gdansk, Poland

N. Calik is with institute of Informatics of Istanbul Medeniyet University, Turkey (e-mail: nurullah.calik@medeniyet.edu.tr).

P. Mahouti is with department of electronic and Automation of Istanbul Cerrahpasa University, Turkey (e-mail: pmahouti@iuc.edu.tr).

Mehmet A. Belen is department of electrical and electronic engineering of Iskenderun Technical University, Turkey (e-mail: mali.belen@iste.edu.tr)

Although the surrogates can be used as stand-alone models, i.e., overall replacements of EM simulations, this is only possible for simple design scenarios (low-dimensional parameter spaces, limited parameter ranges, slightly nonlinear system outputs) [46], [47]. In more realistic situations, data-driven models are often combined with nature-inspired optimization procedures [48]-[50], or machine learning frameworks [51], [52], where utilization of the model as a predictor is interleaved with its updating process, typically using some sort of sequential design of experiments [53] with appropriate infill criteria [54]. The latter can be oriented towards improving the model accuracy [55] or identification of the global optimum [56]. Physics-based surrogates constitute a second group of metamodels, which are particularly popular in the context of local optimization [57]. The main idea is to correct an underlying low-fidelity representation of the system (e.g., equivalent network [58], or coarse-mesh EM model) using sparsely-sampled high-fidelity data. Owing to the embedded problem-specific knowledge, physics-based surrogates exhibit better generalization capability (as compared to data-driven models) but they are less versatile, and highly dependent on the setup (selection of the low-fidelity model and the correction methods). Representative examples of physics-based techniques include space mapping [59], cognition-driven design [60], and various response correction approaches (e.g., [61], [62]).

Despite their popularity, data-driven models are severely affected by the curse of dimensionality, which hinders their application to modeling of real-world antenna structures. Due to high nonlinearity of antenna characteristics, handling more than a few parameters becomes a serious problem, especially when broad ranges thereof are to be covered to ensure design utility of the model. Depending on the problem and setup, the scope of applicability can be somehow extended through dimensionality reduction (principal component analysis [63], model order reduction approaches [64]), high-dimensional model representation (HDMR) [65], orthogonal matching pursuit [66], or variable-fidelity methods (Bayesian model fusion [67], co-kriging [68]). An alternative approach is the employment of deep learning (DL) algorithms [69], in particular, Deep Neural Networks (DNN) that have demonstrated their capability of improved handling of nonlinear system outputs as compared to more traditional regression models (e.g., [70], [71]). Unfortunately, DNNs face some serious issues on their own, primarily related to the model setup (adjustment of hyper-parameters, the network architecture, preventing overtraining, etc.) [72]-[74]. Another alternative is ensemble learning (EL) [75], which is essentially is a technique of strategically combining different sets of individual models, referred to as learners, to solve a specific computer intelligence challenge such as classification or regression. The main principle of EL is to use base learner models, usually exhibiting low performance as stand-along surrogates, as the building blocks for generating surrogates with more complex architecture and improved performance [76]. However, an appropriate handling of the building blocks, including their selection and integration, is a non-trivial task [77]. Recently introduced performance-driven (or constrained) modeling paradigm offers a conceptually different solution to the dimensionality and parameter range issues [78]. Therein, the surrogate model is only constructed in a specified region of the parameter space, containing design that are of high-quality with respect to the assumed performance figures. Identification of this region, and, consequently, the modeling domain, is realized using a set of pre-optimized reference designs and an auxiliary inverse model. The surrogate model itself is constructed using kriging interpolation. As demonstrated in [79] and [80], domain confinement allows for a dramatic improvement of the accuracy while using small number of training data samples.

This work proposes a framework for modeling of antenna structures, which combines the domain confinement approach, and a pyramidal deep regression network (PDRN). The latter is a novel approach incorporating deep neural network surrogates with a pyramidal-shape architecture, and procedures for automated determination of the specific model network and the hyper-parameters. The latter allows for mitigating the common issues pertinent to DNN models, in particular, the risk of overtraining, but also enables improved exploitation of the problem-specific knowledge embedded in the training data set. At the same time, constraining the model domain according to the performance-driven modeling methodology [78], permits efficient handling of parameter space dimensionality and setting up high-quality models valid over broad parameter ranges. The presented procedure has been comprehensively validated using three antenna structures, and benchmarked against conventional surrogates as well as nested kriging [79]. The results demonstrate that our approach fully exploits all advantages implied by domain confinement. At the same time, the predictive power of the obtained models is greatly improved as compared to the nested kriging framework (by a factor of around two on the average), let alone the remaining benchmark methods. The latter is a consequence of inherent suitability of DNN in terms of representing nonlinear frequency characteristics of antenna structures. The proposed approach can be considered a viable alternative to existing techniques whenever a construction of accurate replacement models is required in multi-parameter spaces under limited computational resources.

The novelty and the technical contribution of this work can be summarized as follows: (i) development of a novel ANN-based surrogate retaining the flexibility of deep neural network models but with architecture described by a small number of control parameters, which addresses practical issues pertinent to DNN metamodels (overtraining, etc.); (iii) development of procedures for automated hyper-parameter determination through Bayesian Optimization, (iii) incorporating the proposed surrogate into performance-driven modeling paradigm to alleviate major difficulties of antenna response modeling, (iv) demonstrating the relevance of the proposed approach for modeling of multi-parameter antenna characteristics over broad ranges of geometry parameters and operating conditions, (v) demonstrating superiority of the presented technique over several state-of-the art surrogates, both conventional and domain-confined ones.

II. Antenna Modeling Using Domain-Confined Pyramidal Deep Regression Network

This section introduces the modelling methodology proposed in this work. We start by formulating the antenna modelling task in Section II.*A*. This is followed by an outline of the pyramidal deep regression network (PDRN) surrogate, developed to represent antenna characteristics within the presented framework (Section II.*B*). The concept and implementation of domain confinement [78] is briefly recalled in Section II.*C*, whereas Section II.*D* summarizes the operational flow of the overall modelling procedure. Comprehensive verification of the methodology along with its benchmarking will be discussed in Section III.

A. Surrogate Modeling: Problem Formulation

For an antenna structure of interest, we will denote by $\boldsymbol{x} = [x_1 \ \dots \ x_n]^T$ the vector of its independent parameters. Typically, these are antenna dimensions or material parameters (e.g., permittivity of the substrate, the structure is implemented on). It is assumed that the primary computational model of the antenna is based on full-wave electromagnetic (EM) analysis. The relevant antenna responses (reflection, gain, axial ratio characteristics, etc.) at the design $\boldsymbol{x}$ are collectively denoted as $\boldsymbol{R}(\boldsymbol{x})$. The objective is to construct a fast replacement model (surrogate) $\boldsymbol{R}_s$, which is valid within a certain domain X, typically defined using lower and upper bounds $\boldsymbol{l} = [l_1 \ \dots \ l_n]T$ and $\boldsymbol{u} = [u_1 \ \dots \ u_n]^T$, on the parameters, so that $l_k \le x_k \le u_k$, for $k = 1, \dots, n$. The surrogate should represent the EM model as well as possible over X, i.e., $\boldsymbol{R}(\boldsymbol{x}) - \boldsymbol{R}_s(\boldsymbol{x})$ is to be small (e.g., in the RMS sense [81]) for all $\boldsymbol{x} \in X$. At the same time, the surrogate should be computationally cheap so that multiple evaluations of $\boldsymbol{R}_s$ do not entail considerable CPU expenses.

In this work, we focus on data-driven models, i.e., the surrogates that are constructed using the data pairs $\{\boldsymbol{x}^{(i)}, \boldsymbol{R}(\boldsymbol{x}^{(i)})\}_{i = 1,\dots,N_T}$, acquired across the domain X. This data is approximated, and the model yields predictions concerning the antenna characteristics for arbitrary parameter vectors within X.

B. Pyramidal Deep Regression Network (PDRN) Surrogate

This section introduces the proposed pyramidal deep regression network (PDRN) model. By using a pyramid-like structure of the neuron layers, PDRN attempts to find the optimum model by transforming the input parameters into a very high dimensional space to increase the number of degrees of freedom and facilitate handling of nonlinear input-output relationships pertinent to the antenna structure at hand. Subsequently, after a mapping step, the problem is reduced to a low-dimensional space. For preventing model over-fitting, the model leaky ReLU activation function [82], [83], [84] is employed. The detailed description of the PDRN surrogate is provided below, following a brief characterization of the artificial neural network (ANN) and deep neural network (DNN) models.

Universal approximation is a well-known property of ANN, which indicates a possibility of modelling a continuous function over a compact set [85], [86]. Let $f: \mathbb{R}^{D_i} \to \mathbb{R}^{D_o}$ be a general transformer function over $\mathcal{S} \subset \mathbb{R}^{D_i}$. D_i and D_o stand for the input and output space dimensions. A general feedforward neural network can be defined as

$$f(\boldsymbol{x}) = \sigma(\boldsymbol{W}^K \cdots \sigma(\boldsymbol{W}^2 \times \sigma(\boldsymbol{W}^1 \times \boldsymbol{x} + \boldsymbol{b}_1) \cdots + \boldsymbol{b}_K) \quad (1)$$

where $\boldsymbol{x} \in \mathbb{R}^{D_i \times 1}$, $f(\boldsymbol{x}) \in \mathbb{R}^{D_o \times 1}$ are the input and output vectors, $\boldsymbol{W}^l, \boldsymbol{b}_l$ are the weight matrices and bias vector, respectively, whereas $\sigma(\cdot)$ is a non-linear activation function. Using this definition, it can be seen that $f(\boldsymbol{x})$ is non-linear affine transformation over $\boldsymbol{\mathcal{S}}$. Its non-linearity has a vital role in modelling of $f(\boldsymbol{x})$. In the training phase of neural network, an optimizer attempts to identify the optimal set of weights to minimize a given loss function, defined as a metric of the network approximation capability. ANN models are capable of modelling complex data. ANN enact a significant role in the field of regression studies [41], [42]. It is especially in multi-output problems, where ANN ability to simultaneously produce all outputs at the same time makes it superior to algorithms that can only yield a single output, such as Support Vector Regression Machine (SVRM) [43], or Gaussian Process Regression (GPR) [44]. Otherwise, in these algorithms, separate models must be trained for each output. This situation is troublesome for large data sets data, such as frequency characteristics of antenna structures. While nonlinearity leverages NN structures, it also reveals their weakness. In particular, the asymptotic regions of the activation function determine the characteristics of the neural networks. Sigmoid and tanh, which are the most widely used activation functions in ANN models, can be given as an example. The ANN models learning mechanism is based on portioning the error generated in the loss function onto the weights. The process is referred to as back propagation, and enabled using the derivative operator's chain rule. Let $\mathcal{L}(\boldsymbol{t}, f(\boldsymbol{x}))$ be a loss function for the ANN over $\boldsymbol{t} \in \mathbb{R}^{D_o \times 1}$ target vector. A derivative of this function with respect to $\boldsymbol{W}^l$ can be expressed as

$$\frac{\partial \mathcal{L}(\boldsymbol{t}, f(\boldsymbol{x}))}{\partial \boldsymbol{W}^l} = \frac{\partial \mathcal{L}(\boldsymbol{t}, f(\boldsymbol{x}))}{\partial \sigma^K} \times \frac{\partial \sigma^K}{\partial \boldsymbol{x}^{K-1}} \times \cdots \frac{\partial \sigma^l}{\partial \boldsymbol{W}^l} \quad (2)$$

where σ^l are the outputs of the lth layer activations. Sigmoid and tanh activations have asymptotic regions. Hence, during the back propagation of the error, the derivatives tent to be zero in these regions. This problem is referred to as *vanishing gradient*, where it affects the generalization capabilities of the models using the activation functions featuring a saturation regime [87]. In particular, it makes effective learning of ANN a complicated matter. This issue is especially pronounced for Deep Neural Network (DNN) due to the internal covariate shift problem [88]. Similar problems have been experienced in Convolutional Neural Network (CNN) models, which have been popular in the recent years and have a very deep structure. However, with the increasing interest in Convolutional Neural Network (CNN) structures, many alternative activation functions have been proposed to alleviate this difficulty [82]-[84].

Deep Neural Networks (DNN) is a recent generation of ANN models (Shallow Neural Networks) that generate outputs by passing input information through more than two layers [89]. One of the best known DNN models in this field are CNNs. CNN models have been developed especially for image processing applications. By reducing the raw $N \times M$ sized image to lower dimensions with the convolution mechanism, it

obtains a feature vector and classifies images at the same time. The principal approach here is that the raw image data in very high dimension is gradually mapped to lower dimensions. This way, raw $N \times M$ size data is reduced to D dimension. Inspired by this, we set up a regression model with the strategy of transition from higher dimensional space to lower dimensional space in order to process data featuring complex input/output relationships.

The Pyramidal Deep Regression Network (PDRN) model proposed in this work capitalizes on several improvements and mechanisms introduced into the DNN surrogates. In particular, PDRN is designed to alleviate the issues common to neural networks. The overall architecture of PDRN, shown in Fig. 1, is the pyramid-shaped network. The idea behind this structure is to transform the input parameters into very high dimensional space to increase the number of degrees of freedom and facilitate handling of nonlinear input-output relationships pertinent to the antenna structure at hand. Subsequently, information distillation progresses to low-dimensional space. This process can actually be thought of as a dimensional evolution between the input and output parameters. As mentioned before, PDRN relies on rendering a DNN by reducing the number of neurons for subsequent layers. From the point of view of the network output, the number of neurons is gradually increasing towards the input layer. The scale ratio s_c is fixed for all layers, which allows for simplifying the network architecture by limiting it to a specific (here, trapezoidal) shape, which can be controlled using a few parameters to be automatically determined during the training process.

As seen in the Fig. 1, the neural architecture of PDRN depends of ϕ_K that is the number of neurons for the last layer. The number of neurons for other layers are determined by factoring of ϕ_K. The depth size d_s is another parameter of PDRN, which controls the number of hidden layers between the input and the output layers. For example, according to this definition, $\phi_K = 64$, $s_c = 1.2$ and $d_s = 3$ yields the architecture of PDRN as

$$\mathcal{M}: r\left(64\times(sc)^2\right)\rightarrow r\left(64\times(sc)^1\right)\rightarrow r\left(64\times(sc)^0\right) \qquad (3)$$

where $\mathcal{M}$ represents the model structure and $r(\cdot)$ is rounding operator to ensure and integer number of neurons in each layer. Thus, the PDRN structure can be fully controlled using the three parameters, ϕ_K, s_c, and d_s. As mentioned before, the classical activation function lead to the vanishing gradient issues in DNN. To prevent this phenomena, a leaky ReLU (*LReLU*) activation employed in CNNs is selected [90] to work with PDRN. LReLU activation and its derivative are defined as

$$\sigma_\alpha(x)=\begin{cases} x & x\geq 0 \\ \alpha x & otherwise \end{cases} \qquad \frac{d\sigma_\alpha(x)}{dx}=\begin{cases} 1 & x\geq 0 \\ \alpha & otherwise \end{cases} \qquad (4)$$

where the parameter $\alpha \in (0,1)$ is referred to as the leakage ratio. Leaky ReLU has a negative slope, and this makes it operable at every gradient flow. As can be observed that the derivative of the activation function is non-zero in every training iteration. This makes it suitable for modelling complex data relations. In PDRN, LReLU activation and α are fixed for all layers, therefore, it is sufficient to only select one parameter (α) in the construction phase of the surrogate. In summary, the PDRN hyper-parameters determining the architecture of the model are: (i) the number of neurons in the last layer ϕ_K, (ii) the depth of the model d_s, (iii) the scale parameter for determination of neurons in other layers s_c, and (iv) the leakage parameter α of the LReLU.

Let $\boldsymbol{v} \in \mathbb{R}^{D_i\times 1}$ be the input vector containing the parameters of the regression problem. PDRN maps it to a desired output parameter domain. The output values are compared with the target values using the Mean Absolute Error (MAE) loss function, and the error is propagated to the neural weights. The MAE loss is defined as,

$$\mathcal{L}_{MAE}(\boldsymbol{o}) = \mathbb{E}\{|\boldsymbol{o}-\boldsymbol{t}|\} \qquad (5)$$

where $\boldsymbol{o}, \boldsymbol{t} \in \mathbb{R}^{D_o\times 1}$ stand for the prediction values of network and the target vector in D_o dimension output space, respectively. $\mathbb{E}\{\cdot\}$ is expectation operator. This loss indicates that the training process is equivalent to the ℓ_1-norm minimization task.

In DNN models, including the proposed PDRN surrogate, it is vital to use a pre-processing step for heterogeneous-valued data which may feature considerably different parameter ranges. For data standardization, a z-score methodology is preferred [91], [92]. Let $\boldsymbol{V} \in \mathbb{R}^{N_{tr}\times D_i}$ stores the training samples as row vector. Z-score is defined as,

$$\boldsymbol{v}_i^z = \frac{1}{N_{tr}}\times\sum_i \frac{\boldsymbol{v}_i-\boldsymbol{\mu}_z}{\boldsymbol{\eta}_z} \qquad (6)$$

where $\boldsymbol{\mu}_z, \boldsymbol{\eta}_z$ and $\boldsymbol{v}_i^z$ are sample mean, standard deviation of dimensions of $\boldsymbol{V}$, and normalized training sample. Division in (6) is understood in the Hadamard sense (component-wise).

Having established the generic architecture with hyper-parameters along with the pre-processed training data, the next step is the model training process itself. The key element thereof is to compute the derivatives of the loss function with respect to the network weights. By using gradient-based descent algorithms, the model is trained to attain a local optimum in the solution space.

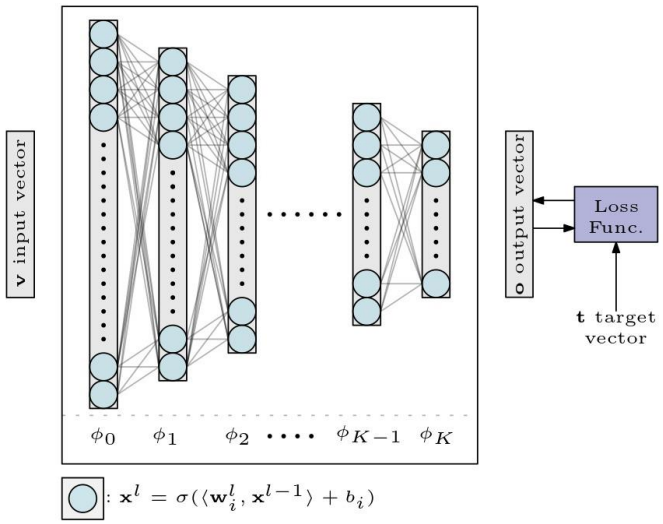

Fig. 1. Generic PDRN model architecture: $\boldsymbol{x}^l$ and $\boldsymbol{w}^l$ are output and weight values of lth layer. $\sigma(\cdot)$ is the activation function.

If the DNN models are trained with a small amount of data, the training error can be readily reduced with no improvement of the model predictive power. This situation is referred to as overfitting [87]. Utilization of large-scale training data prevents the model from entering the overfitting regime. Although handling large-scale data is problematic for algorithms such as SVRM or GPR, it can be tackled with *stochastic gradient descent*-based optimizers such as Adagrad [93], RMSProp [94], and Adam [95]. Stochastic gradient-based descent procedures are the preferred training strategies for neural networks, which differ from a classical approach in utilization of the error function gradients. The former uses the gradients to train the model after passing the entire data through the network. Stochastic gradient descent uses a partition of data called mini batch to generate gradients, then train neurons by using this particular subset. Sweeping the entire data determines one training epoch. In this study, Adam optimizer is preferred to update the model weights. One of the fundamental issues in DNN, including the proposed model, is the adjustment of the network architecture, which is often realized by trial and error. In this work, Bayesian Optimization (BO) is employed to determine the architecture-defining hyper-parameters of PDRN in a fully automated manner. This will be elaborated on in Section II.*C*.

C. PDRN Architecture Search via Bayesian Optimization

Optimal hyper-parameters determination is critical for the models designed for any classification or regression tasks where these parameters can be continuous variables (such as weights) or discrete variables such as the number of layers and the number of neurons. Herein, the optimal hyper-parameter determination of the proposed PDRN model is executed by means of Bayesian Optimization (BO). BO utilizes probabilistic modelling techniques to enable global optimisation of complex and expensive functions [96].

Probabilistic Surrogate Model (PSM) of the objective function iteratively improved at each observation, and the acquisition function (used to explore the state space based on the surrogate model optimally) are the two essential parts of Bayesian Optimization (BO) [97], [98], [99]. While PSM is being used to evaluate the uncertainty of the model, the acquisition function allows for balancing the exploration and exploitation during the optimization run. The model uncertainty plays a key role in this process. One of the most commonly used acquisition function is Expected Improvement (EI) [100]. The initial models are generated based on the random sampling or design of experiment (DoE) [101]. At each iteration, the PSMs are updated using the sampled data. Subsequent (in-fill) sample points are then generated by maximizing the acquisition function. Table I provides a pseudocode of the BO.

Although various models can be used for the prior distribution of probabilistic model [101], Gaussian Process (GP) is applied more frequently for BO as compared to others [102], [103]. For $f: X \to R$, GP assumes that for any finite set of samples $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\} \in X^n$, the vector $(f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_n))^T$ follows joint multivariate Gaussian distribution

$$\left[f(\boldsymbol{x}_1) \dots f(\boldsymbol{x}_n) \right]^T \sim N(\boldsymbol{\mu}, K) \tag{7}$$

where $\boldsymbol{\mu}$ is a $n \times 1$ mean vector, K is $n \times n$ a covariance matrix.

TABLE I
PSEUDOCODE OF BAYESIAN OPTIMIZATION

Bayesian Optimization[#] [96]

1: Initial Sampling
2: Generate probabilistic surrogate model
3: f**or** t = 1, 2, 3, ….. **do**
4: Find x_t optimizing the acquisition function
5: Sample $\boldsymbol{y}_i = f(\boldsymbol{x}_i) + N(0, \sigma^2_{noise})$,
6: update the probabilistic model
7: End for
8: return optimal $\boldsymbol{f}(\boldsymbol{x})$

[#] Here, x_i is the ith sample, y_i is the noisy observation of the objective function at x_i.

GP can be defined by the mean function $m(\boldsymbol{x})$ and the covariance function $k(\boldsymbol{x}_1, \boldsymbol{x}_2)$. The mean vector μ is determined as

$$\mu_i = m(\boldsymbol{x}_i), \ \forall i \in 1, \dots, n \tag{8}$$

whereas the covariance matrix is determined by

$$K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j) \tag{9}$$

Herein, Squared Exponential Covariance (Kernel) is used as the covariance function

$$k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma_f^2 \exp\left[-\frac{1}{2} \frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^T (\boldsymbol{x}_i - \boldsymbol{x}_j)}{\sigma_l^2} \right] \tag{10}$$

where σ_l is the characteristic length scale, and σ_f is the standard deviation of the signal. Let $D = \{X, \boldsymbol{y}\}$ be the training set, where $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ and $\boldsymbol{y} = \{f(x_1), \dots, f(x_n)\}$. For a new observation point $\boldsymbol{x}_*$, the corresponding function value $\boldsymbol{y}_* = f(\boldsymbol{x}_*)$ follows joint Gaussian distribution

$$\begin{bmatrix} \boldsymbol{y}_* \\ \boldsymbol{y} \end{bmatrix} \sim N\left(\begin{bmatrix} m(\boldsymbol{x}_*) \\ m \end{bmatrix}, \begin{bmatrix} k(\boldsymbol{x}_*, \boldsymbol{x}_*) & k(\boldsymbol{x}_*, X) \\ k(X, \boldsymbol{x}_*) & K_n \end{bmatrix} \right) \tag{11}$$

where $\boldsymbol{m} = \left(m(\boldsymbol{x}_1), \dots, m(\boldsymbol{x}_n)\right)^T$ is the $n \times 1$ mean vector, l is the length scale, and

$$k(\boldsymbol{x}_*, X) = (k(\boldsymbol{x}_*, \boldsymbol{x}_1), \dots, k(\boldsymbol{x}_*, \boldsymbol{x}_n)) \tag{12}$$

$$k(X, \boldsymbol{x}_*) = k(\boldsymbol{x}_*, X)^T \tag{13}$$

$$K_n = \begin{bmatrix} k(\boldsymbol{x}_1, \boldsymbol{x}_1) & \dots & k(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ k(\boldsymbol{x}_2, \boldsymbol{x}_1) & \dots & k(\boldsymbol{x}_2, \boldsymbol{x}_n) \\ \vdots & \ddots & \vdots \\ k(\boldsymbol{x}_n, \boldsymbol{x}_1) & \dots & k(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix} \tag{14}$$

Here, it should be noted that y_* is bounded by $\boldsymbol{y}$ and also follows the normal distribution [104] defined as

$$\boldsymbol{y}_* \sim N(\mu_{\boldsymbol{y}_*|\boldsymbol{y}}, \sigma^2_{\boldsymbol{y}_*|\boldsymbol{y}}) \tag{15}$$

$$\mu_{\boldsymbol{y}_*|\boldsymbol{y}} = m(\boldsymbol{x}_*) + k(\boldsymbol{x}_*, X) K_n^{-1} (\boldsymbol{y} - \boldsymbol{m}) \tag{16}$$

$$\sigma^2_{\boldsymbol{y}_*|\boldsymbol{y}} = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, X) K_n^{-1} k(X, \boldsymbol{x}_*) \tag{17}$$

The mean $\mu_{\boldsymbol{y}_*|\boldsymbol{y}}$ can be taken as the prediction while $\sigma^2_{\boldsymbol{y}_*|\boldsymbol{y}}$ is the confidence of the prediction for probabilistic prediction of $\boldsymbol{y}_*$. A Gaussian noise $\epsilon \sim N(0, \sigma_n^2)$, σ_n^2: variance of Gaussian noise, can be added to $f(\boldsymbol{x})$ for observation of the noise effects. The covariance function and prediction of GP model after applying noise would be defined as:

$$k_n(\boldsymbol{x}_i, \boldsymbol{x}_j) = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \sigma_n^2 \delta_{ij} \qquad \delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \tag{18}$$

$$\mu_{\boldsymbol{y}_*|\boldsymbol{y}} = m(\boldsymbol{x}_*) + k(\boldsymbol{x}_*, X)(K_n + \sigma_n^2 \delta_{ij})^{-1} (\boldsymbol{y} - \boldsymbol{m}) \tag{19}$$

$$\sigma^2_{y_*|y} = k(\boldsymbol{x}_*, \boldsymbol{x}_*) - k(\boldsymbol{x}_*, X)(K_n + \sigma_n^2 \delta_{ij})^{-1} k(X, \boldsymbol{x}_*) \quad (20)$$

The acquisition function is a key component of BO, which is used to determine the new (infill) points by balancing the trade-off between exploration and exploitation [98]. By using GP, the acquisition function attempts to minimize the objective function $\boldsymbol{y} = f(\boldsymbol{x})$. As mentioned before, one of the commonly used acquisition functions is Expected Improvement (EI) [105], [106]. Let take τ as the currently known minimal value of $f(\boldsymbol{x})$. The improvement I for y can be defined as

$$I(y,\tau) = \begin{cases} \tau - y & y < \tau \\ 0 & otherwise \end{cases} \quad (21)$$

and indicates that the improvement only occur when y is less than τ. Based on the GP, the prediction of *y* at observation point *x* use the distribution of

$$p(\boldsymbol{y}|D, \mathcal{S}) = N(\mu(\boldsymbol{x}), \sigma^2(\boldsymbol{x})) \quad (22)$$

where $\mu(x)$ and σ are the mean and variance of the distribution, $\mathcal{S}$ is the vector of hyper-parameters. Based on these, EI can be defined as

$$\begin{aligned} EI(\boldsymbol{x}) &= E\left[I(\boldsymbol{y},\tau)\right] = \int_{-\infty}^{\infty} I(\boldsymbol{y},\tau)\, p(\boldsymbol{y}|D,\mathcal{S})\, d\boldsymbol{y} \\ &= (\tau - \mu(\boldsymbol{x}))\,\Omega\left(\frac{\tau - \mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right) + \sigma(\boldsymbol{x})\,\omega\left(\frac{\tau - \mu(\boldsymbol{x})}{\sigma(\boldsymbol{x})}\right) \end{aligned} \quad (23)$$

where $\Omega(\cdot)$ is the cumulative distribution function of standard normal distribution, and $\omega(\cdot)$ is the probability density function.

In the following, the application of BO for PDRN neural architecture search is presented. BO estimates the number of neurons in the last layer ϕ_K), the depth d_s of the model, the scale parameter s_p, and the leakage parameter α of LReLU. The PDRN architecture parameters can be jointly described as $S = \{\phi_K, d_s, s_c, \alpha\}$. The operation of the BO search algorithm in the context of PDRN has been summarized in Table II.

Figure 2 shows the flow chart of the PDRN neural architecture search by BO. The output of the algorithm is the set of optimal hyper-parameters, which corresponds to the minimum average *k*-fold loss value $\mathcal{L}_{(avg)}$ [107-110]. BO randomly generated a sample from the given hyper-parameter space that would be used to evaluate the *k*-fold error for a model created using these parameters. Subsequently, the parameter set $\mathcal{S}$, and the average loss value $\mathcal{L}_{(avg)}$ are stored, and $\mathcal{S}$ is updated. Towards the end of the optimization process, the parameter set $\mathcal{S}$ that gives the minimum average loss value $\mathcal{L}_{(avg)}$ will be taken as the optimal model for the given training dataset.

In this study, the parameter ranges for the neural architecture search using BO are taken as follows: $\phi_K \in \{16, 32, 48, \ldots, 128\}$, $d_s \in \{3, 4, 5, 6, 7\}$, $s_c \in \{1, 1.25, 1.5, \ldots, 3\}$, $\alpha \in [0, 0.5]$. Here, in order to reduce the computational overhead of the optimization process, ϕ_K and s_c are taken as discrete values. Although these value can also be considered as continues variables, the effect of small variations in their values can generally be neglected. The maximum number of iteration is taken as 30 in this study. As mentioned before, over-fitting is one of the most challenging problems of DNN. Apart from using the appropriate activation function, the employment of regularization layers such as Dropout and Batch normalizations are efficient solutions to prevent the over-fitting of the models [111], [112].

Construction of the proposed PDRN model is a three-stage process as explained in Fig. 3. In Stage 1, BO is used to determine the optimum network architecture for a given training data set. In this work, k = 2-fold is selected to evaluate the neural architecture quality. In Stage 2, the best hyper-parameter set $\mathcal{S}^{best}$ is used to generate the PDRN model, which is then trained using the entire input data set. In Stage 3, the final model performance is evaluated using the test data.

TABLE II
BAYESIAN OPTIMIZATION OF PDRN HYPER-PARAMETERS

Algorithm 2: Bayesian Optimization Search Algorithm

Input of $\mathcal{S}(.)$: ***V***, *iter*
Output $\{\phi_K, d_s, s_c, \alpha\} = \mathcal{S}$ ($\boldsymbol{V}_{tr}$, *iter*)

Procedure

1. Set $t = 0$; Initialize $S^t = \{\phi_K, d_s, s_c, \alpha\}$
2. Generate PDRN model by using S^t
3. Evaluate *k*-fold average score *Lavg* for the model
4. Update *S* using BO, store $\{S^t, L_{avg}\}$, and set $t = t + 1$
5. **If** t = *iter*, **then** terminate the process, **else** go to 3
6. **Return** S^{best} corresponding to minimum L_{avg}

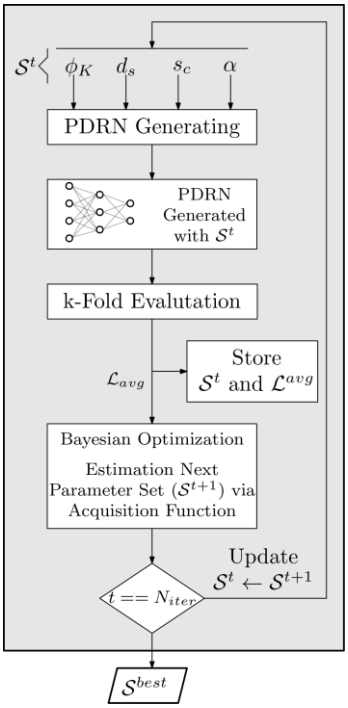

Fig. 2. Flow diagram of the neural architecture search of PDRN using Bayesian optimization.

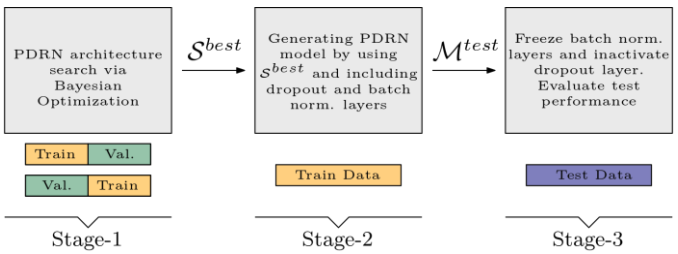

Fig. 3. Flow diagram of the PDRN surrogate training and architecture optimization.

D. Domain Confinement

As it will be demonstrated in Section 3, PDRN does ensure improved handling of highly-nonlinear antenna characteristics as compared to standard regression surrogates, e.g., kriging or support vector regression. In this work, for further enhancement of the accuracy, it is coupled with the performance-driven modeling paradigm of [79]. For the convenience of the reader, this section briefly outlines the concept and analytical formulation of domain confinement.

The key concept is the objective space F that consists of objective vectors $\boldsymbol{f} = [f_1 \ \ldots \ f_N]^T$, where f_k is the kth figure of interest (e.g., center frequency, bandwidth, power split ratio, substrate permittivity, etc.) or material parameter (e.g., permittivity of the substrate the antenna is to be implemented on). The space F is delimited by the lower and upper bounds $f_{k.\min}$ and $f_{k.\max}$, $k = 1, \ldots, N$, that determine the intended region of validity of the surrogate model.

The second fundamental concept of performance-driven modeling is design optimality. Given the objective function $U(\boldsymbol{x},\boldsymbol{f})$ that quantifies the quality of the design $\boldsymbol{x}$ for a selected objective vector $\boldsymbol{f}$, the optimum design $\boldsymbol{x}^*$ is understood as

$$\boldsymbol{x}^* = U_F(\boldsymbol{f}) = \arg\min_{\boldsymbol{x}} U(\boldsymbol{x}, \boldsymbol{f}) \tag{24}$$

As an elementary example, consider a patch antenna to be matched at the center frequency f_0, and implemented on the dielectric substrate of relative permittivity ε_r. The surrogate model of the antenna input characteristic is to be constructed within the range of center frequencies $f_{0.\min} \le f_0 \le f_{0.\max}$, and the range of permittivity $\varepsilon_{r.\min} \le \varepsilon_r \le \varepsilon_{r.\max}$. Then, the objective space F is the interval $[f_{0.\min}\ f_{0.\max}] \times [\varepsilon_{r.\min}\ \varepsilon_{r.\max}]$, and the objective function may be defined as $U(\boldsymbol{x},\boldsymbol{f}) = \max\{|S_{11.\varepsilon r}(\boldsymbol{x},f_0)|\}$, where $S_{11.\varepsilon r}(\boldsymbol{x},f)$ is the antenna reflection coefficient at the design $\boldsymbol{x}$ and frequency f, simulated for the structure implemented on the substrate with permittivity ε_r.

The set of all designs optimal in the sense of (24) for all objective vectors $\boldsymbol{f} \in F$ will be denoted as

$$U_F(F) = \{U_F(\boldsymbol{f}) : \boldsymbol{f} \in F\} \tag{25}$$

It can be observed that restricting the surrogate model domain to $U_F(F)$ and its vicinity is sufficient to capture all designs that are of interest in a particular design context (i.e., from the point of view of the objective f_k) [79]. Thus, the immediate task is to approximate $U_F(F)$ as a subset of the parameter space X. In the nested kriging framework [79] this is realized using a set of reference designs $\boldsymbol{x}^{(j)} = [x_1^{(j)} \ \ldots \ x_n^{(j)}]^T$ $j = 1, \ldots, p$, pre-optimized w.r.t. the vectors $\boldsymbol{f}^{(j)} = [f_1^{(j)} \ \ldots \ f_N^{(j)}]$, which are distributed (in a possibly uniform manner) in F. In practice, the reference designs have to be acquired beforehand, but in some cases, they may be available from the previous design work with the same antenna structure.

The definition of the constrained domain of the surrogate model is a two-step process. In the first step, the first-level kriging interpolation model $\boldsymbol{s}_I(\boldsymbol{f}) : F \to X$ is obtained using the training data set $\{\boldsymbol{f}^{(j)},\boldsymbol{x}^{(j)}\}$, $j = 1, \ldots, p$. Note that $\boldsymbol{s}_I$ is an inverse model with $\boldsymbol{s}_I(F)$ approximating $U_F(F)$. Furthermore, we have $\boldsymbol{s}_I(\boldsymbol{f}^{(j)}) = \boldsymbol{x}^{(j)}$ for $j = 1, \ldots, p$, but the two sets do not coincide outside the reference set. This makes it necessary to fatten $\boldsymbol{s}_I(F)$ so that the entire $U_F(F)$ can be accommodated therein. In [76], this is realized by extending $\boldsymbol{s}_I(F)$ using the orthonormal basis $\{\boldsymbol{v}_n^{(k)}(\boldsymbol{f})\}$, $k = 1, \ldots, n - N$, of vectors normal to $\boldsymbol{s}_I(F)$ at $\boldsymbol{f}$. To this end, the extension coefficients are defined as

$$\begin{aligned}\boldsymbol{\alpha}(\boldsymbol{f}) &= [\alpha_1(\boldsymbol{f}) \ldots \alpha_{n-N}(\boldsymbol{f})]^T = \\ &= 0.5T\left[|\,\boldsymbol{x}_d \boldsymbol{v}_n^{(1)}(\boldsymbol{f})\,| \ \ldots \ |\,\boldsymbol{x}_d \boldsymbol{v}_n^{(n-N)}(\boldsymbol{f})\,|\right]^T\end{aligned} \tag{26}$$

where $\boldsymbol{x}_d = \boldsymbol{x}_{\max} - \boldsymbol{x}_{\min}$, with $\boldsymbol{x}_{\max} = \max\{\boldsymbol{x}^{(k)}, k = 1, \ldots, p\}$, $\boldsymbol{x}_{\min} = \min\{\boldsymbol{x}^{(k)}, k = 1, \ldots, p\}$. Here, T is a thickness parameter. The vertical bounds of the surrogate model domain X_S are the manifolds M_+ and M_-

$$M_\pm = \left\{\boldsymbol{x} \in X : \boldsymbol{x} = \boldsymbol{s}_I(\boldsymbol{f}) \pm \sum_{k=1}^{n-N} \alpha_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f})\right\} \tag{27}$$

The domain itself is then defined as [76]

$$X_S = \left\{ \begin{array}{c} \boldsymbol{x} = \boldsymbol{s}_I(\boldsymbol{f}) + \sum_{k=1}^{n-N} \lambda_k \alpha_k(\boldsymbol{f}) \boldsymbol{v}_n^{(k)}(\boldsymbol{f}) : \boldsymbol{f} \in F, \\ -1 \le \lambda_k \le 1, k = 1, \ldots, n-N \end{array} \right\} \tag{28}$$

Figure 4 provides a graphical illustration of the concepts involves in the domain definition. The second-level (ultimate) surrogate is rendered in X_S usin the training data $\{\boldsymbol{x}_B^{(k)},\boldsymbol{R}(\boldsymbol{x}_B^{(k)})\}_{k = 1, \ldots, NB}$. The data samples $\boldsymbol{x}_B^{(k)}$ are uniformly allocated in X_S [79].

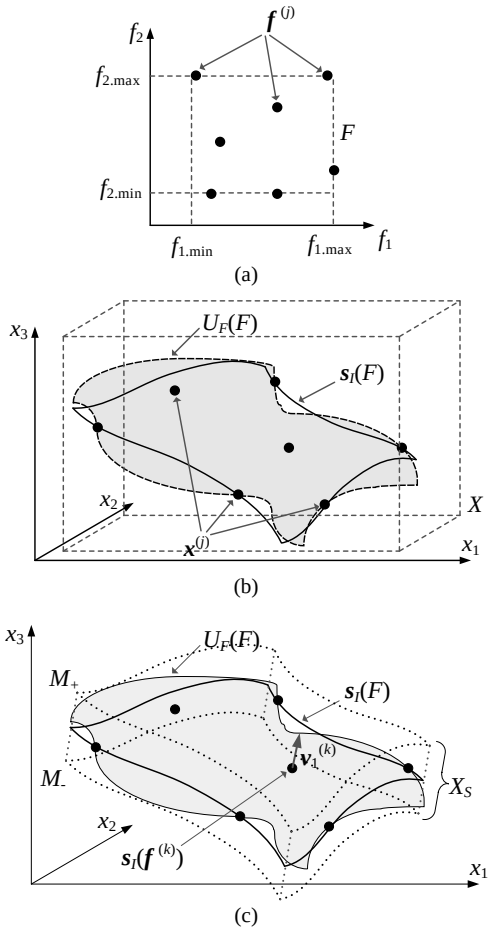

Fig. 4. Performance-driven (constrained) modeling: (a) graphical illustration of the objective space (here shown for two performance figures); (b) parameter space X, reference designs, and the optimum design set $U_F(F)$ along with the image $\boldsymbol{s}_I(F)$ of the first-level surrogate; note that $U_F(F)$ and $\boldsymbol{s}_I(F)$ only coincide at the reference designs, therefore, extension is necessary; (c) definition of the domain X_S through orthogonal extension of $\boldsymbol{s}_I(F)$: for illustration the normal vector $\boldsymbol{v}_1^{(k)}$ is shown at $\boldsymbol{f}^{(k)}$ along with the manifolds M_- and M_+ and the domain X_S.

As expected, domain confinement has a profound effect on the model predictive power. In particular, it opens the door to constructing accurate metamodels over broad ranges of antenna geometry parameters and operating conditions using reasonably small numbers of training samples [79]. This feature is explored in the modeling methodology proposed in this work.

E. Modeling Framework

Figure 5 summarizes the operational flow of the modeling methodology proposed in this work. The objective space is set up by the user according to the intended range of validity of the surrogate. The reference designs are assumed to be acquired beforehand. The first step of the procedure is to define the model domain, which is followed by design of experiments (cf. [79] for details), and the acquisition of the EM simulation data. The PDRN surrogate is then established by first determining its specific architecture, then the model itself (hyper-parameters, etc.).

III. Verification Case Studies and Benchmarking

This section provides a comprehensive numerical validation of the modelling methodology proposed in this work. On the one hand, we investigate the advantages of modelling antenna characteristics using PDRN rather than standard regression techniques. On the other hand, we demonstrate the benefits of combining PDRN with the domain confinement paradigm. The surrogates are constructed using the data sets of various sizes, from 50 to 800 samples. Furthermore, constrained PDRN is compared to several benchmark methods, including kriging, support vector regression, and ensemble learning. The experiments are based on modelling of reflection and gain characteristics of three microstrip antennas with the figures of interest including operating frequencies as well as substrate permittivity.

A. Test Cases

The geometries of the verification antenna structures, referred to as Antenna I, II, and III, are shown in Fig. 6. Antenna I is a uniplanar dual-band dipole [113]. Antenna II is a ring-slot structure [114], whereas Antenna III is a quasi-Yagi structure with a parabolic reflector [115]. Table 3 provides the details concerning the antenna substrates and geometry parameters as well as the computational models. It should be noted that for Antennas II and III, the relative permittivity of the substrate is to be considered as one of the operating conditions (i.e., a component of the objective space), so that the surrogate model to be constructed is valid for a range of permittivity values.

B. Experimental Setup. Benchmark Methods

Table IV gathers the information concerning the parameter and objective spaces, as well as the reference designs. In all cases, the objective spaces are two-dimensional. For Antenna I, the operating conditions are the operating frequencies of the lower and upper operating band. In the case of Antennas II and III, the first component of the objective space is the operating frequency, whereas the second is the relative permittivity of the dielectric substrate the antenna is to be implemented on. The objective spaces in Table IV determine the intended region of validity of the surrogate to be constructed. The allocation of the reference designs can be found in the second-but-last row of the table; the average number of designs is ten. The parameter spaces are determined as the smallest intervals containing the reference points. It can be observed that the parameter ranges are wide, which, along with broad ranges of the operating conditions, make the modeling tasks extremely challenging. The details concerning design optimality are provided in the last row of Table IV.

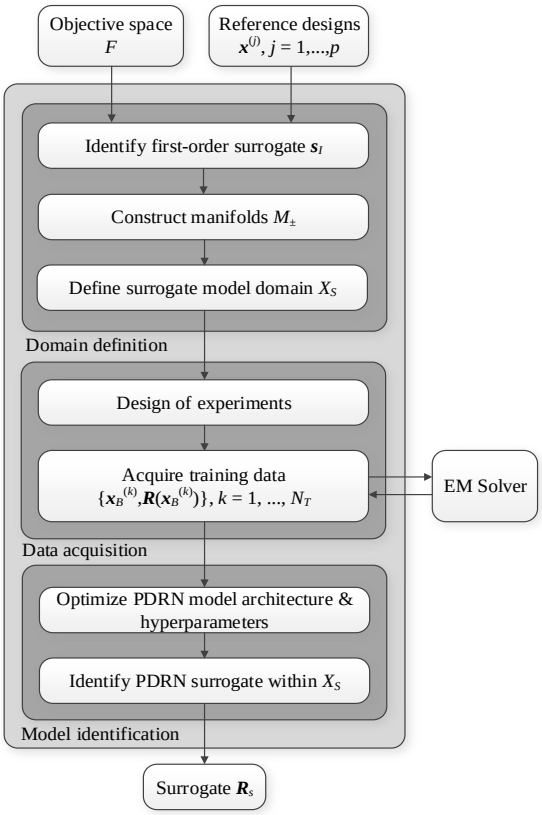

Fig. 5. Flowchart of the proposed modeling framework involving PDRN with automated architecture determination and domain confinement.

TABLE III
Verification Case Studies: Antenna Parameters and Computational Models

	Antenna		
	I	II	III
Substrate	RO4350 (ε_r = 3.5, h = 0.76 mm)	h = 0.76 mm ε_r – one of objective space components	h = 1.5 mm ε_r – one of objective space components
Independent parameters	$\boldsymbol{x} = [l_1\ l_2\ l_3\ w_1\ w_2\ w_3]^T$	$\boldsymbol{x} = [l_f\ l_d\ w_d\ r\ s\ s_d\ o\ g]^T$	$\boldsymbol{x} = [W\ L\ L_m\ L_p\ S_d\ S_r\ W_2\ W_a\ W_d\ g]^T$
Other parameters	l_0 = 30, w_0 = 3, s_0 = 0.15, and o = 5	w_f computed for given ε_r to ensure 50-ohm line impedance	W_1 computed for given ε_r to ensure 50-ohm line impedance
Computational model	CST (~100,000 cells, simulation time, ~60 seconds)	CST (~300,000 cells, simulation time, ~90 seconds)	CST (~1,100,000 cells, simulation time, ~3 minutes)

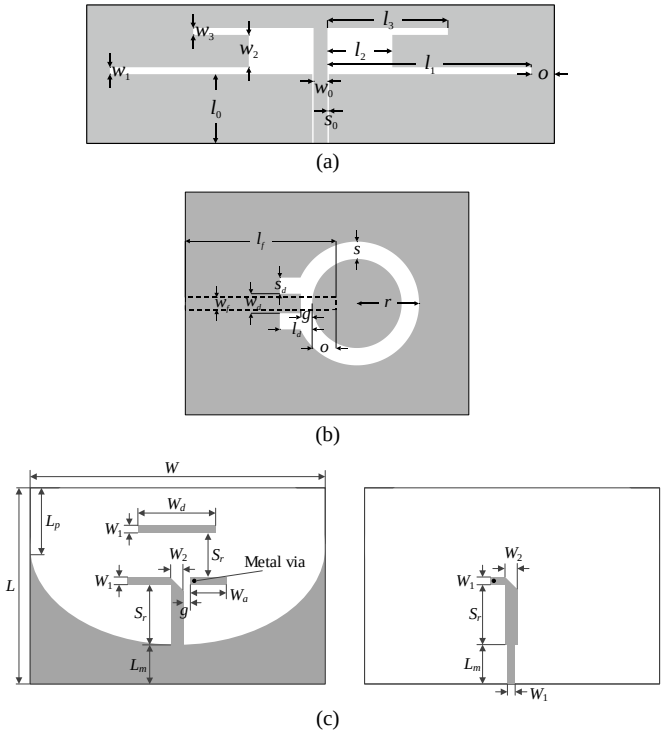

Fig. 6. Geometries of the verification antenna structures: (a) Antenna I [114], (b) Antenna II [114], and (c) Antenna III (front and back on the left- and right-hand side of the picture) [115].

For all test antennas, the surrogate models are constructed both in the original parameter space X, and in the confined domain defined as described in Section II.*C*, using the thickness parameters $T = 0.05$ in all cases. The models are rendered using five different data sets containing 50, 100, 200, 400, and 800 samples. This is to investigate the scalability of the models, i.e., the relationship between the dataset cardinality and the model predictive power.

The modeling error is estimated using a split-sample method based on a separate set of 100 testing samples. The error measure is the relative RMS error defined as $\|\boldsymbol{R}(\boldsymbol{x}) - \boldsymbol{R}_s(\boldsymbol{x})\|/\|\boldsymbol{R}(\boldsymbol{x})\|$, where $\boldsymbol{R}$ and $\boldsymbol{R}_s$ stand for the EM simulated and surrogate-evaluated antenna characteristics (complex in the case of reflection response, and dB-valued for realized gain).

The following four techniques are considered:

- The modeling approach proposed in this work, pyramidal deep regression network (PDRN) of Section 2.2;
- Kriging interpolation [55];
- Convolutional neural network (CNN) [43];
- Ensemble learning [75];

The last three methods are the benchmark approaches. The kriging surrogate used second-order polynomial (as a trend function) and Gaussian correlation function [55]. CNN algorithm had been used as a counterpart method [82-84], CNN had been used as a state of the art regression method for prediction of reflection phase characteristics of a reflect-array FSS in [116], same hyper parameters (Depth size of 3 and first filter amount of 64) in [116] are also used here for generating a counterpart model for proposed PDRN. Finally, the ensemble learning method had been used as a final counterpart model. The ensemble aggregation method had been taken as LSBoost, which is one of the regression ensembles that aims at minimizing the mean-squared error [117], Learning rate of the model is obtained via the use of Bayesian Optimization [77] using the data set belong to Antenna I in parameter space X with 50 samples which is found as 0.24.

C. Results and Discussion

The numerical results have been gathered in Tables V through VIII. Figures 5, 6, and 7 show the responses of the proposed surrogate model and the EM-simulation data for selected test designs.

TABLE IV
VERIFICATION CASE STUDIES: PARAMETER AND OBJECTIVE SPACES, REFERENCE DESIGNS

	Antenna		
	I	II	III
Parameter space X (lower and upper bounds $\boldsymbol{l}$ and $\boldsymbol{u}$)	$\boldsymbol{l} = [29\ 5.0\ 17\ 0.2\ 1.5\ 0.5]^T$ $\boldsymbol{u} = [42\ 12\ 25\ 0.6\ 5.2\ 3.5]^T$	$\boldsymbol{l} = [22.0\ 3.5\ 0.3\ 6.5\ 3.0\ 0.5\ 3.5\ 0.2]^T$ $\boldsymbol{u} = [27.0\ 8.0\ 2.3\ 16.0\ 7.0\ 5.5\ 6.0\ 2.3]^T$	$\boldsymbol{l} = [100\ 55\ 10\ 14.5\ 6.0\ 10\ 2.0\ 7.5\ 16.3\ 0.5]^T$ $\boldsymbol{u} = [137\ 81\ 29\ 28\ 21\ 18\ 5.0\ 20\ 40\ 1.0]^T$
Objective space F	Lower and upper operating frequencies: 2.0 GHz $\le f_1 \le$ 3.0 GHz; 4.0 GHz $\le f_2 \le$ 5.5 GHz => $\boldsymbol{f} = [f_1\ f_2]^T$	Operating frequency: 2.5 GHz $\le f_0 \le$ 6.5 GHz; Substrate permittivity $2.0 \le \varepsilon_r \le 5.0$ => $\boldsymbol{f} = [f_0\ \varepsilon_r]^T$	Operating frequency: 2.5 GHz $\le f_0 \le$ 5.0 GHz; Substrate permittivity $2.5 \le \varepsilon_r \le 4.5$ => $\boldsymbol{f} = [f_0\ \varepsilon_r]^T$
Reference designs[$]	$\{f_1,f_2\}$ = {2.0, 4.0}, {2.2, 5.0}, {2.0, 5.5}, {2.3, 4.5}, {2.4, 5.5}, {2.6, 4.0}, {2.7, 3.5}, {2.8, 4 .7}, {3.0, 4.0}, and {3.0, 3.5}.	$\{f,\varepsilon_r\}$ = {2.5,2.0}, {2.5,3.5}, {2.5,5.0}, {4.0,3.5}, {4.5,2.0}, {4.5,5.0}, {5.0,3.5}, {6.5,2.0}, {6.5,3.5}, {6.5,5.0}.	$\{f_0,\varepsilon_r\}$ = {2.5, 4.5}, {3.5, 4.5}, {5.0, 4.5}, {2.5, 2.5}, {5.0, 2.5}, {3.5, 2.5}, {4.5, 3.5}, and {3.0, 3.5}
Design optimality[#]	$U(\boldsymbol{x},\boldsymbol{f}) = \max\{\lvert S_{11}(\boldsymbol{x},f_1)\rvert, \lvert S_{11}(\boldsymbol{x},f_2)\rvert\}$	$U(\boldsymbol{x},\boldsymbol{f}) = \max\{\lvert S_{11}(\boldsymbol{x},f_0)\rvert\}$	$U(\boldsymbol{x},\boldsymbol{f}) = U(\boldsymbol{x},[f_0\ \varepsilon_r]^T) =$ $-(f_0B)^{-1}\int_{f_0(1-B/2)}^{f_0(1+B/2)} G(\boldsymbol{x},f)df + \beta c(\boldsymbol{x})$ Here, $G(\boldsymbol{x},f)$ is the realized gain at frequency f; $\beta c(\boldsymbol{x})$ is a penalty term

[$]The allocation of the reference designs in terms of the pairs of the objective vector components.

[#]Formulation of the objective function defining the design optimality. Antenna I and II are to be optimized in a minimax sense. Antenna III is to be designed for maximum (average) in-band gain while ensuring –10 dB in-band matching. The bandwidth B is 0.08 (eight percent); β is a penalty coefficient and c is a penalty function that quantifies violation of the constraint $|S_{11}| \le -10$ dB (e.g., $c(\boldsymbol{x}) = \max\{\max\{f_0(1 - B/2) \le f \le f_0(1 + B/2) : |S_{11}(\boldsymbol{x},f)| + 10\},0\}^2$).

TABLE V
MODELING RESULTS AND BENCHMARKING FOR ANTENNA I

Number of training samples	Relative RMS error							
	Modeling in parameter space X				Modeling in confined domain X_S			
	Kriging	CNN	Ensemble learning	Proposed PDRN	Kriging	CNN	Ensemble learning	Proposed PDRN
50	21.7 %	29.2 %	31.1 %	10.4 %	6.9 %	8.6 %	14.5 %	3.0 %
100	17.3 %	16.1 %	29.1 %	7.0 %	4.5 %	6.0 %	13.1 %	2.4 %
200	12.6 %	10.2 %	27.5 %	4.1 %	2.8 %	4.5 %	11.6 %	1.9 %
400	9.3 %	7.0 %	25.7 %	3.3 %	2.6 %	3.9 %	11.0 %	1.6 %
800	7.2 %	6.1 %	25.0 %	2.7 %	2.4 %	3.7 %	10.7 %	1.6 %

TABLE VI
MODELING RESULTS AND BENCHMARKING FOR ANTENNA II

Number of training samples	Relative RMS error							
	Modeling in parameter space X				Modeling in confined domain X_S			
	Kriging	CNN	Ensemble learning	Proposed PDRN	Kriging	CNN	Ensemble learning	Proposed PDRN
50	56.9 %	67.7 %	73.8 %	36.2 %	12.9 %	13.1 %	28.9 %	4.1 %
100	50.8 %	58.8 %	69.1 %	23.3 %	6.9 %	8.2 %	25.4 %	2.6 %
200	35.8 %	34.0 %	63.9 %	12.8 %	4.9 %	6.4 %	23.1 %	2.3 %
400	31.5 %	22.3 %	58.1 %	7.5 %	3.1 %	4.8 %	19.9 %	1.9 %
800	25.6 %	13.5 %	55.8 %	5.3 %	2.2 %	5.3 %	19.1 %	1.5 %

TABLE VII
MODELING RESULTS AND BENCHMARKING FOR ANTENNA III

Number of training samples	Relative RMS error							
	Modeling in parameter space X				Modeling in confined domain X_S			
	Kriging	CNN	Ensemble learning	Proposed PDRN	Kriging	CNN	Ensemble learning	Proposed PDRN
50	61.4 %	70.1 %	56.4 %	30.7 %	17.9 %	31.3 %	34.1 %	9.7 %
100	50.7 %	56.4 %	54.5 %	19.6 %	13.3 %	21.5 %	27.1 %	7.7 %
200	39.8 %	46.6 %	46.1 %	13.8 %	7.5 %	15.7 %	24.5 %	5.3 %
400	32.8 %	38.0 %	44.4 %	10.7 %	5.4 %	11.9 %	22.7 %	3.9 %
800	31.8 %	29.8 %	40.4 %	8.8 %	4.5 %	9.3 %	19.9 %	3.2 %

TABLE VIII
OPTIMALLY SELECTED PDRN MODEL ARCHITECTURES

Case Study		Antenna I				Antenna II				Antenna III ($\|S_{11}\|$)				Antenna III (Gain)			
Number of training samples	Domain	ϕ_K	d_s	s_p	α	ϕ_K	d_s	s_p	α	ϕ_K	d_s	s_p	α	ϕ_K	d_s	s_p	α
50	original X	48	5	2.5	0.13	48	6	1.75	0.03	96	6	1.5	0.06	80	5	1.25	0.18
	confined X_S	64	6	2	0.06	64	5	2.5	0.03	80	6	1.75	0.01	48	5	2	0.05
100	original X	32	6	2.5	0.02	32	6	2.5	0.04	80	6	1.5	0.06	80	5	1.5	0.09
	confined X_S	64	6	1.5	0.17	64	6	2	0.02	96	6	1.25	0.18	96	6	2	0.01
200	original X	64	5	2.25	0.02	32	6	1.5	0.01	80	6	1.75	0.03	96	6	2	0.03
	confined X_S	64	5	2.5	0.01	48	6	1.25	0.01	64	5	2	0.02	80	6	1.5	0.02
400	original X	64	5	2.5	0.01	64	6	2	0.02	32	6	2	0.02	96	4	2	0.05
	confined X_S	48	6	1.75	0.03	48	5	2	0.01	96	5	1.25	0.06	80	6	2	0.001
800	original X	48	6	2.5	0.01	48	5	1.75	0.01	64	4	2	0.08	32	6	1.5	0.05
	confined X_S	48	6	1.25	0.01	64	6	1.5	0.01	96	4	2	0.01	48	6	1.75	0.01

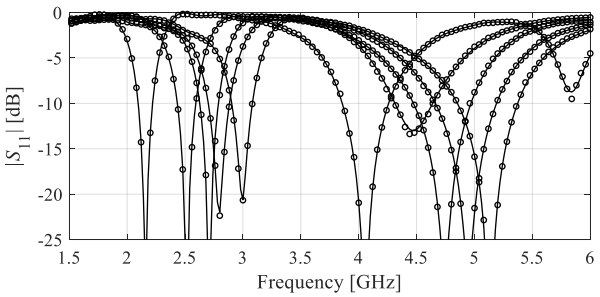

Fig. 5. Antenna I: responses of the proposed surrogate model constructed in the confined domain using 400 training samples (o) versus EM-simulation data (—). The curves represent antenna input characteristics as the selected test designs.

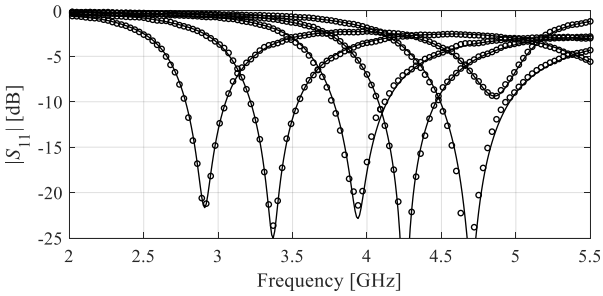

Fig. 6. Antenna II: responses of the proposed surrogate model constructed in the confined domain using 400 training samples (o) versus EM-simulation data (—). The curves represent antenna input characteristics as the selected test designs.

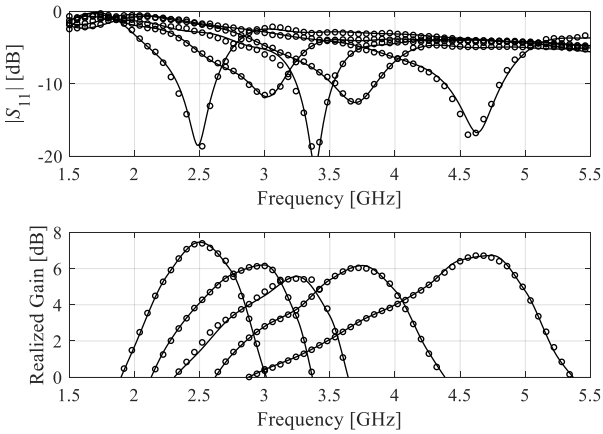

Fig. 7. Antenna III: responses of the proposed surrogate model constructed in the confined domain using 400 training samples (o) versus EM-simulation data (—). The curves represent antenna input characteristics (top) and realized gain (bottom) as the selected test designs.

The hyper-parameters of the proposed PDRN surrogate have been optimized on Intel Core i7-76000K CPU @ 4 GHz with 16 GB RAM and GTX1080TI GPU. The computing times of model training were as follows:

- Antenna I: 3, 7, 18, 39, and 91 minutes for 50, 100, 200, 400, and 800 samples, respectively (space X); 3, 7, 19, 43, and 96 minutes for 50, 100, 200, 400, and 800 samples, respectively (confined domain X_S);
- Antenna II: 5, 11, 25, 55, and 127 minutes for 50, 100, 200, 400, and 800 samples, respectively (space X); 6, 13, 26, 59, and 134 minutes for 50, 100, 200, 400, and 800 samples, respectively (confined domain X_S);
- Antenna III: 5, 13, 31, 69, and 157 minutes for 50, 100, 200, 400, and 800 samples, respectively (space X); 9, 14, 34, 74, and 162 minutes for 50, 100, 200, 400, and 800 samples, respectively (confined domain X_S).

It can be observed that the computational complexity of the model training process grows slightly faster than linearly with respect to the training data set size.

The data provided in the Tables V through VIII allows for drawing conclusions concerning the performance of the proposed modeling approach in both the original and confined spaces. These are summarized below:

- The proposed PDRN model exhibits a considerably better predictive power than the benchmark techniques, both in the original parameter space X and in the confined domain X_S.
- The differences in the modeling error are significant in favor of PDRN, and they are more pronounced for smaller training sets (50 to 200 samples), which indicates that the proposed technique is capable of exploiting the allocation of the available data through the automated architecture and hyper-parameter determination;
- The above observation is confirmed in Table VIII, which indicates that the specific model architecture achieved during the training phase depends on both the considered antenna structure and the training data set size;
- PDRN model capitalizes on domain confinement, which allows for further improvements of the predictive power of the surrogate, especially for smaller data sets;
- In particular, constrained PDRN outperforms the nested kriging framework and the accuracy improvement is by a factor two to three for the data sets of 50 samples, and by about 1.5 for the largest sets. These differences are significant given that all considered test cases are very challenging both in terms of the parameter space dimensionality and—even more importantly—the parameter and operating condition ranges;
- The obtained results are consistent throughout the entire set of antenna structures utilized for verification, which demonstrates that the proposed technique can be successfully used for handling a variety of antenna modeling tasks without any additional tuning of its control parameters.

Overall, to the authors' knowledge, the performance of the proposed constrained PDRN surrogate is perhaps the best ever reported in the antenna modeling literature thus far. It should also be mentioned that experimental validation of the presented antennas is not included in this work as being irrelevant to the paper topic. Notwithstanding, it can be found in other works employing these antenna structures for benchmark purposes (e.g., [79], [81]).

IV. Conclusion

This paper proposed a novel framework for accurate and reduced-cost modelling of antenna structures. Our methodology combines enhanced pyramidal deep regression network (PDRN) surrogates with automated architecture and hyper-parameter determination, and the concept of performance-driven modelling. The former offers improved flexibility in terms of representing highly-nonlinear antenna characteristics, whereas the latter enables significant

computational savings in terms of training data acquisition. The presented approach has been demonstrated using three microstrip antennas with their replacement models constructed over broad ranges of geometry and material parameters as well as operating frequencies. All considered test cases are challenging also due to the dimensionality of the parameter spaces as well as the parameter ranges. Comprehensive benchmarking indicates superiority of the domain-confined PDRN technique over both conventional regression models (kriging, CNN) but also the ensemble learning model. In particular, the predictive power of the confined surrogates has been improved by a factor of two over the nested kriging method without increasing cardinality of the training data set. The proposed framework can be suitable for rendering reliable surrogates under challenging scenarios, such modelling of multi-band antennas over wide ranges of centre frequencies, or modelling of multiple frequency characteristics of antennas implemented on various substrates of different height and permittivity.

ACKNOWLEDGMENT

The authors would like to thank Dassault Systemes, France, for making CST Microwave Studio available. This work is partially supported by the Icelandic Centre for Research (RANNIS) Grant 217771 and by National Science Centre of Poland Grant 2018/31/B/ST7/02369.

REFERENCES

[1] S. Fu, Z. Cao, X. Quan, and C. Xu, "A broadband dual-polarized notched-band antenna for 2/3/4/5G base station," *IEEE Ant. Wireless Prop. Lett.*, vol. 19, no. 1, pp. 69-73, 2020 doi: 10.1109/LAWP.2019.2953294.

[2] X. Cai and K. Sarabandi, "Broadband omnidirectional circularly polarized antenna with asymmetric power divider," *IEEE Transactions Ant. Propag.*, vol. 68, no. 7, pp. 5171-5181, 2020, doi: 10.1109/TAP.2020.2977725.

[3] M. Kovaleva, D. Bulger, and K. P. Esselle, "Comparative study of optimization algorithms on the design of broadband antennas," *IEEE J. Multiscale Multiphysics Comp. Techn.*, vol. 5, pp. 89-98, 2020, doi: 10.1109/JMMCT.2020.3000563.

[4] J. Ren, W. Hu, Y. Yin, and R. Fan, "Compact printed MIMO antenna for UWB applications," *IEEE Ant. Wireless Prop. Lett.*, vol. 13, pp. 1517-1520, 2014, doi: 10.1109/LAWP.2014.2343454.

[5] G. Teni, N. Zhang, J. Qiu, and P. Zhang, "Research on a novel miniaturized antipodal vivaldi antenna with improved radiation," *IEEE Ant. Wireless Prop. Lett.*, vol. 12, pp. 417-420, 2013, doi: 10.1109/LAWP.2013.2253592.

[6] J. Tao and Q. Feng, "Compact ultrawideband MIMO antenna with half-slot structure," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 792-795, 2017, doi: 10.1109/LAWP.2016.2604344.

[7] B. R. S. Reddy and D. Vakula, "Compact zigzag-shaped-slit microstrip antenna with circular defected ground structure for wireless applications," *IEEE Ant. Wireless Prop. Lett.*, vol. 14, pp. 678-681, 2015, doi: 10.1109/LAWP.2014.2376984.

[8] J. Liu, Z. Tang, Z. Wang, H. Li, and Y. Yin, "Gain enhancement of a broadband symmetrical dual-loop antenna using shorting pins," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 8, pp. 1369-1372, 2018, doi: 10.1109/LAWP.2018.2844293.

[9] U. Ullah, I. B. Mabrouk, and S. Koziel, "A compact circularly polarized antenna with directional pattern for wearable off-body communications," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 12, pp. 2523-2527, 2019, doi: 10.1109/LAWP.2019.2942147.

[10] Z. Wang, C. Li, Q. Wu, and Y. Yin, "A metasurface-based low-profile array decoupling technology to enhance isolation in MIMO antenna systems," *IEEE Access*, vol. 8, pp. 125565-125575, 2020, doi: 10.1109/ACCESS.2020.3007188.

[11] L. Kurra, M. P. Abegaonkar, A. Basu, and S. K. Koul, "FSS properties of a uniplanar EBG and its application in directivity enhancement of a microstrip antenna," *IEEE Ant. Wireless Prop. Lett.*, vol. 15, pp. 1606-1609, 2016, doi: 10.1109/LAWP.2016.2518299.

[12] D. Gao, Z. Cao, X. Quan, M. Sun, S. Fu, and P. Chen, "A low-profile decoupling slot-strip array for 2 × 2 microstrip antenna," *IEEE Access*, vol. 8, pp. 113532-113542, 2020, doi: 10.1109/ACCESS.2020.3002862.

[13] M.A. Haq and S. Koziel, "On topology modifications for wideband antenna miniaturization," *AEU - Int. J. Electr. Comm.*, vol. 94, pp. 215-220, 2018, doi: 10.1016/j.aeue.2018.07.006.

[14] P. Zhou, Z. Zhang, and M. He, "Radiation pattern recovery of the impaired-radome-enclosed antenna array," *IEEE Ant. Wireless Prop. Lett.*, vol. 19, no. 9, pp. 1639-1643, 2020, doi: 10.1109/LAWP.2020.3013228.

[15] S. Su, C. Lee, and S. Chen, "Very-low-profile, triband, two-antenna system for WLAN notebook computers," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 9, pp. 1626-1629, 2018, doi: 10.1109/LAWP.2018.2858849.

[16] H. Geeta, G. Patil, and Dinesh, "Design and study of EM coupled dual band linear array antenna with point to multipoint communication for millimeter wave application," *Int. Conf. Electrical, Electronics, Communication, Computer, and Opt. Techn. (ICEECCOT)*, Msyuru, India, 2018, pp. 1599-1603, 2018, doi: 10.1109/ICEECCOT43722.2018.9001617.

[17] U. Ullah, S. Koziel, and I. B. Mabrouk, "Rapid redesign and bandwidth/size trade-offs for compact wideband circular polarization antennas using inverse surrogates and fast EM-based parameter tuning," *IEEE Trans. Ant. Prop.*, vol. 68, no. 1, pp. 81-89, 2019, doi: 10.1109/TAP.2019.2935817.

[18] D. De and P. K. Sahu, "Design of an endfire microstrip antenna for aircraft collision avoidance system," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 5, pp. 996-1000, 2019.

[19] S. Kim and S. Nam, "Compact ultrawideband antenna on folded ground plane," *IEEE Trans. Ant. Prop.*, vol. 68, no. 10, pp. 7179-7183, 2020, doi: 10.1109/LAWP.2019.2907477.

[20] S. Lei, Y. Yang, H. Hu, Z. Zhao, B. Chen, and X. Qiu, "Power gain optimization method for wide-beam array antenna via convex optimization," *IEEE Trans. Ant. Prop.*, vol. 67, no. 3, pp. 1620-1629, 2019, doi: 10.1109/TAP.2018.2884814.

[21] Q. Li, Q. Chu, Y. Chang, and J. Dong, "Tri-objective compact log-periodic dipole array antenna design using MOEA/D-GPSO," *IEEE Trans. Ant. Prop.*, vol. 68, no. 4, pp. 2714-2723, 2020, doi: 10.1109/TAP.2019.2949705.

[22] J. Du and C. Roblin, "Statistical modeling of disturbed antennas based on the polynomial chaos expansion," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 1843-1846, 2017, doi: 10.1109/LAWP.2016.2609739.

[23] A. Pietrenko-Dabrowska, S. Koziel, and M. Al-Hasan, "Expedited yield optimization of narrow- and multi-band antennas using performance-driven surrogates," *IEEE Access*, vol. 8, pp. 143104-143113, 2020, doi: 10.1109/ACCESS.2020.3013985.

[24] J. Wang, X. S. Yang, and B. Z. Wang, "Efficient gradient-based optimization of pixel antenna with large-scale connections," *IET Microwaves Ant. Prop.*, vol. 12, no. 3, pp. 385-389, 2018, doi: 10.1049/iet-map.2017.0719.

[25] S. Koziel and A. Pietrenko-Dabrowska, "Reduced-cost electromagnetic-driven optimization of antenna structures by means of trust-region gradient-search with sparse Jacobian updates," *IET Microwaves Ant. Prop.*, vol. 13, no. 10, pp. 1646-1652, 2019, doi: 10.1049/iet-map.2018.5879.

[26] S. Koziel and A. Pietrenko-Dabrowska, "Variable-fidelity simulation models and sparse gradient updates for cost-efficient optimization of compact antenna input characteristics," *Sensors*, vol. 19, no. 8, 2019, doi: 10.3390/s19081806.

[27] F. Arndt, "WASP-NET®: Recent advances in fast full 3D EM CAD of waveguide feeds and aperture antennas," *IEEE Int. Symp. Ant. Propag., APS-URSI*, Spokane, WA, pp. 2724-2727, 2011, doi: 10.1109/APS.2011.5997088.

[28] B. B. Tierney and A. Grbic, "Designing anisotropic, inhomogeneous metamaterial devices through optimization," *IEEE Trans. Ant. Prop.*, vol. 67, no. 2, pp. 998-1009, 2019, doi: 10.1109/TAP.2018.2883668.

[29] A. S. O. Hassan, A. S. Etman, and E. A. Soliman, "Optimization of a novel nano antenna with two radiation modes using kriging surrogate models," *IEEE Photonic J.*, vol. 10, no. 4, art. no. 4800807, 2018, doi: 10.1109/JPHOT.2018.2848593.

[30] J. Du and C. Roblin, "Stochastic surrogate models of deformable antennas based on vector spherical harmonics and polynomial chaos

expansions: application to textile antennas," *IEEE Trans. Antennas Propag.*, vol. 66, no. 7, pp. 3610-3622, 2018, doi: 10.1109/TAP.2018.2829820.

[31] S. Koziel and S. Ogurtsov, "Antenna design by simulation-driven optimization. Surrogate-based approach," *Springer*, New York, 2014.

[32] C. Wang, S. Yuan, W. Gao, C. Jiang, C. Zhu, P. Li, Z. Wang, X. Peng, and Y. Shi, "A Taylor-surrogate-model-based method for the electrical performance of array antennas under interval position errors," *IEEE Ant. Wireless Prop. Lett.*, vol. 19, no. 7, pp. 1221-1225, 2020, doi: 10.1109/LAWP.2020.2995873.

[33] H. Yang, F. Lu, H. Guo, and R. Liu, "Design of a new N-shape composite ultra-thin deployable boom in the post-buckling range using response surface method and optimization," *IEEE Access*, vol. 7, pp. 129659-129665, 2019, doi: 10.1109/ACCESS.2019.2934744.

[34] S. Koziel and S. Ogurtsov, Simulation-based optimization of antenna arrays, *World Scientific*, 2019, doi: 10.1142/q0179.

[35] D. R. Prado, J. A. López Fernández, M. Arrebola, M. R. Pino, and G. Goussetis, "General framework for the efficient optimization of reflectarray antennas for contoured beam space applications," *IEEE Access*, vol. 6, pp. 72295-72310, 2018, doi: 10.1109/ACCESS.2018.2882271.

[36] T. N. Kapetanakis, I. O. Vardiambasis, M. P. Ioannidou, and A. Maras, "Neural network modeling for the solution of the inverse loop antenna radiation problem," *IEEE Trans. Ant. Prop.*, vol. 66, no. 11, pp. 6283-6290, 2018, doi: 10.1109/TAP.2018.2869136.

[37] J. P. Jacobs, "Characterization by Gaussian processes of finite substrate size effects on gain patterns of microstrip antennas," *IET Microwaves Ant. Prop.*, vol. 10, no. 11, pp. 1189-1195, 2016, doi: 10.1049/iet-map.2015.0621.

[38] J. L. Chávez-Hurtado and J. E. Rayas-Sánchez, "Polynomial-based surrogate modeling of RF and microwave circuits in frequency domain exploiting the multinomial theorem," *IEEE Trans. Microwave Theory and Tech.*, vol. 64, no. 12, pp. 4371-4381, 2016, doi: 10.1109/TMTT.2016.2623902.

[39] F. Declercq, I. Couckuyt, H. Rogier, and T. Dhaene, "Environmental high frequency characterization of fabrics based on a novel surrogate modelling antenna technique," *IEEE Trans. Ant. Prop.*, vol. 61, no. 10, pp. 5200-5213, 2013, doi: 10.1109/TAP.2013.2274031.

[40] P. Barmuta, F. Ferranti, G. P. Gibiino, A. Lewandowski and D. M. M. P. Schreurs, "Compact behavioral models of nonlinear active devices using response surface methodology," *IEEE Trans. Microwave Theory and Tech.*, vol. 63, no. 1, pp. 56-64, 2015, doi: 10.1109/TMTT.2014.2376559.

[41] J. Dong, W. Qin, and M. Wang, "Fast multi-objective optimization of multi-parameter antenna structures based on improved BPNN surrogate model," *IEEE Access*, vol. 7, pp. 77692-77701, 2019, doi: 10.1109/ACCESS.2019.2920945.

[42] L. Yuan, X. Yang, C. Wang, and B. Wang, "Multibranch artificial neural network modeling for inverse estimation of antenna array directivity," *IEEE Trans. Ant. Prop.*, vol. 68, no. 6, pp. 4417-4427, 2020, doi: 10.1109/TAP.2020.2970071.

[43] D. R. Prado, J. A. López-Fernández, M. Arrebola, and G. Goussetis, "Support vector regression to accelerate design and crosspolar optimization of shaped-beam reflectarray antennas for space applications," *IEEE Trans. Ant. Prop.*, vol. 67, no. 3, pp. 1659-1668, 2019, doi: 10.1109/TAP.2018.2889029.

[44] J. Gao, Y. Tian, and X. Chen, "Antenna optimization based on co-training algorithm of gaussian process and support vector machine," *IEEE Access*, vol. 8, pp. 211380-211390, 2020, doi: 10.1109/ACCESS.2020.3039269.

[45] A. Petrocchi, A. Kaintura, G. Avolio, D. Spina, T. Dhaene, A. Raffo, and D.M.P.-P. Schreurs, "Measurement uncertainty propagation in transistor model parameters via polynomial chaos expansion," *IEEE Microwave Wireless Comp. Lett.*, vol. 27, no. 6, pp. 572-574, 2017, doi: 10.1109/lmwc.2017.2701334.

[46] J. A. Easum, J. Nagar, and D. H. Werner, "Multi-objective surrogate-assisted optimization applied to patch antenna design," *Int. Symp. Ant. Prop.*, pp. 339-340, San Diego, USA, 2017, doi: 10.1109/apusncursinrsm.2017.8072212.

[47] D. I. L. de Villiers, I. Couckuyt, and T. Dhaene, "Multi-objective optimization of reflector antennas using kriging and probability of improvement," *Int. Symp. Ant. Prop.*, pp. 985-986, San Diego, USA, 2017, doi: 10.1109/apusncursinrsm.2017.8072535.

[48] D. K. Lim, D. K. Woo, H. K. Yeo, S. Y. Jung, J. S. Ro, and H. K. Jung, "A novel surrogate-assisted multi-objective optimization algorithm for an electromagnetic machine design," *IEEE Trans. Magn.*, vol. 51, no. 3, 2015, Art no. 8200804, doi: 10.1109/tmag.2014.2359452.

[49] B. Xia, Z. Ren, and C. S. Koh, "Utilizing kriging surrogate models for multi-objective robust optimization of electromagnetic devices," *IEEE Trans. Magn.*, vol. 50, no. 2, Feb 2014, Art no. 7017104, doi: 10.1109/TMAG.2013.2284925.

[50] A. Toktas, D. Ustun, and M. Tekbas, "Multi-objective design of multi-layer radar absorber using surrogate-based optimization," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 8, pp. 3318-3329, 2019, doi: 10.1109/TMTT.2019.2922600.

[51] Q. Wu, H. Wang, and W. Hong, "Multistage collaborative machine learning and its application to antenna modeling and optimization," *IEEE Trans. Ant. Prop.*, vol. 68, no. 5, pp. 3397-3409, 2020, doi: 10.1109/TAP.2019.2963570.

[52] A. M. Alzahed, S. M. Mikki, and Y. M. M. Antar, "Nonlinear mutual coupling compensation operator design using a novel electromagnetic machine learning paradigm," *IEEE Ant. Wireless Prop. Lett.*, vol. 18, no. 5, pp. 861-865, 2019, doi: 10.1109/LAWP.2019.2903787.

[53] H. M. Torun and M. Swaminathan, "High-dimensional global optimization method for high-frequency electronic design," *IEEE Trans. Microwave Theory Techn.*, vol. 67, no. 6, pp. 2128-2142, 2019, doi: 10.1109/TMTT.2019.2915298.

[54] B. Liu, S. Koziel, and N. Ali, "SADEA-II: a generalized method for efficient global optimization of antenna design," *J. Comp. Design Eng.*, vol. 4, no. 2, pp. 86-97, 2017, doi: 10.1016/j.jcde.2016.11.002.

[55] A. I. J. Forrester and A. J. Keane, "Recent advances in surrogate-based optimization," *Prog. Aerospace Sci.*, vol. 45, pp. 50-79, 2009, doi: 10.1016/j.paerosci.2008.11.001.

[56] S. Xiao, G. Q. Liu, K. L. Zhang, Y. Z. Jing, J. H. Duan, P. Di Barba, and J. K. Sykulski, "Multi-objective Pareto optimization of electromagnetic devices exploiting kriging with Lipschitzian optimized expected improvement," *IEEE Trans. Magn.*, vol. 54, no. 3, 7001704, 2018, doi: 10.1109/TMAG.2017.2771561.

[57] J. C. Cervantes-González, J. E. Rayas-Sánchez, C. A. López, J. R. Camacho-Pérez, Z. Brito-Brito, and J. L. Chávez-Hurtado, "Space mapping optimization of handset antennas considering EM effects of mobile phone components and human body," *Int. J. RF Microwave CAE*, vol. 26, no. 2, pp. 121–128, 2016, doi: 10.1002/mmce.20945.

[58] J. W. Bandler, Q. S. Cheng, S. A. Dakroury, A. S. Mohamed, M. H. Bakr, K. Madsen, and J. Sondergaard, "Space mapping: the state of the art," *IEEE Trans. Microwave Theory Tech.*, vol. 52, no. 1, pp. 337-361, 2004, doi: 10.1109/TMTT.2003.820904.

[59] I. A. Baratta, C. B. de Andrade, R. R. de Assis, and E. J. Silva, "Infinitesimal dipole model using space mapping optimization for antenna placement," *IEEE Ant. Wireless Prop. Lett.*, vol. 17, no. 1, pp. 17-20, 2018, doi: 10.1109/LAWP.2017.2771721.

[60] C. Zhang, F. Feng, V. Gongal-Reddy, Q. J. Zhang, and J. W. Bandler, "Cognition-driven formulation of space mapping for equal-ripple optimization of microwave filters," in *IEEE Trans. Microwave Theory Techn.*, vol. 63, no. 7, pp. 2154-2165, 2015, doi: 10.1109/TMTT.2015.2431675.

[61] S. Koziel and L. Leifsson, "Simulation-driven design by knowledge-based response correction techniques," S*pringer*, New York, 2016.

[62] S. Koziel and S. D. Unnsteinsson "Expedited design closure of antennas by means of trust-region-based adaptive response scaling," *IEEE Antennas Wireless Prop. Lett.*, vol. 17, no. 6, pp. 1099-1103, 2018.

[63] J.A. Tomasson, S. Koziel, and A. Pietrenko-Dabrowska, "Quasi-global optimization of antenna structures using principal components and affine subspace-spanned surrogates," *IEEE Access*, vol. 8, no. 1, pp. 50078-50084, 2020, doi: 10.1109/ACCESS.2020.2980057.

[64] Y. Sato, F. Campelo, and H. Igarashi, "Fast shape optimization of antennas using model order reduction," *IEEE Trans. Magn.*, vol. 51, no. 3, pp. 1-4, Art no. 7204304, 2015, doi: 10.1109/TMAG.2014.2358295.

[65] A. C. Yücel, H. Bağcı, and E. Michielssen, "An ME-PC enhanced HDMR method for efficient statistical analysis of multiconductor transmission line networks," *IEEE Trans. Comp. Packaging and Manufacturing Techn.*, vol .5, no. 5, pp. 685-696, 2015, doi: 10.1109/TCPMT.2015.2424679.

[66] X. Li, "Finding deterministic solution from underdetermined equation: large-scale performance modeling of analog/RF circuits," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems* (*TCAD*), vol. 29, no. 11, pp. 1661-1668, 2010, doi: 10.1109/TCAD.2010.2061292.

[67] S. Koziel, S. Ogurtsov, I. Couckuyt, and T. Dhaene, "Variable-fidelity electromagnetic simulations and co-kriging for accurate modeling

of antennas," *IEEE Trans. Antennas Prop.*, vol. 61, no. 3, pp. 1301-1308, 2013, doi: 10.1109/TAP.2012.2231924.

[68] F. Wang, P. Cachecho, W. Zhang, S. Sun, X. Li, R. Kanj and C. Gu, "Bayesian model fusion: large-scale performance modeling of analog and mixed-signal circuits by reusing early-stage data," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 35, no. 8, pp. 1255-1268, 2016, doi: 10.1145/2463209.2488812.

[69] J. H. Kim and S. W. Choi, "A deep learning-based approach for radiation pattern synthesis of an array antenna," *IEEE Access*, vol. 8, pp. 226059-226063, 2020, doi: 10.1109/ACCESS.2020.3045464.

[70] J. A. Baker and J. P. Jacobs, "Empirical investigation of benefits of increased neural network depth for modeling of antenna input characteristics," *Int. Conf. Electromagnetics in Adv. Appl. (ICEAA)*, Granada, Spain, pp. 1180-1181, 2019, doi: 10.1109/ICEAA.2019.8879115.

[71] C. P. Schwegmann, W. Kleynhans, B. P. Salmon, L. W. Mdakane, and R. G. V. Meyer, "Very deep learning for ship discrimination in synthetic aperture radar imagery," *in 2016 IEEE Trans. Geosci. and Remote Sens. Symposium (IGARSS)*, 2016, pp. 104–107, doi: 10.1109/IGARSS.2016.7729017.

[72] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no .1, pp. 281-305, 2012.

[73] P. Neary, "Automatic hyperparameter tuning in deep convolutional neural networks using asynchronous reinforcement learning," *IEEE Int. Conf. Cognitive Computing (ICCC)*, San Francisco, CA, USA, pp. 73-77, 2018, doi: 10.1109/ICCC.2018.00017.

[74] X. Y. Chen, X. Y. Peng, Y. Peng, and J.-B. Li, "The classification of synthetic aperture radar image target based on deep learning," *J. Inf. Hiding Multimedia Signal Process.*, vol. 7, pp. 1345–1353, 2016.

[75] R. Polikar, "Ensemble learning," *In Ensemble machine learning*, pp. 1-34. Springer, Boston, MA, 2012.

[76] J. M. Moreira, C. Soares, A. M. Jorge, and J. F. D. Sousa, "Ensemble approaches for regression: a survey," *ACM Computing Surveys*, vol. 45, no. 1, pp. 1-40, 2012, doi: 10.1145/2379776.2379786.

[77] J. C. Lévesque, C. Gagné, and R. Sabourin, "Bayesian hyperparameter optimization for ensemble learning," *arXiv preprint arXiv:1605.06394*, 2016.

[78] S. Koziel, "Low-cost data-driven surrogate modeling of antenna structures by constrained sampling," *IEEE Antennas Wireless Prop. Lett.*, vol. 16, pp. 461-464, 2017, doi: 10.1109/LAWP.2016.2583474.

[79] S. Koziel and A. Pietrenko-Dabrowska, "Performance-based nested surrogate modeling of antenna input characteristics," *IEEE Trans. Ant. Prop.*, vol. 67, no. 5, pp. 2904-2912, 2019, doi: 10.1109/TAP.2019.2896761.

[80] S. Koziel and A.T. Sigurdsson, "Triangulation-based constrained surrogate modeling of antennas," *IEEE Trans. Ant. Prop.*, vol. 66, no. 8, pp. 4170-4179, 2018, doi: 10.1109/TAP.2018.283975

[81] S. Koziel and A. Pietrenko-Dabrowska, "Performance-driven surrogate modeling of high-frequency structures," *Springer*, New York, 2020.

[82] H. Zhang, T. W. Weng, P. Y. Chen, C. J. Hsieh, and L. Daniel, "Efficient neural network robustness certification with general activation functions," *arXiv preprint arXiv:1811.00866.*

[83] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," *In Proceedings of the fourteenth international conference on artificial intelligence and statistics*, JMLR Workshop and Conference Proceedings, vol. 5, pp.315-323, 2011.

[84] V. Nair, and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines", *International Conference on Machine Learning*, Haifa, Israel, 2010.

[85] G. C. Calafiore, S. Gaubert, and C. Possieri. "A universal approximation result for difference of log-sum-exp neural networks," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 12, pp. 5603-5612, Dec. 2020, doi: 10.1109/TNNLS.2020.2975051.

[86] F. Fan, J. Xiong, and G. Wang, "Universal approximation with quadratic deep networks," *Neural Networks*, 124, 383-392, 2020, doi: 10.1016/j.neunet.2020.01.007.

[87] A. Rakitianskaia, and A. Engelbrecht, "Measuring saturation in neural networks," *2015 IEEE Symposium Series on Computational Intelligence*, Cape Town, South Africa, 2015, pp. 1423-1430, doi: 10.1109/SSCI.2015.202.

[88] S. Ioffe, and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *In International conference on machine learning*, PMLR, pp. 448-456, 2015.

[89] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters," *in IEEE Trans. Microw. Theory Tech.*, vol. 67, no. 10, pp. 4140- 4155, Oct. 2019, doi: 10.1109/TMTT.2019.2932738.

[90] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint* arXiv:1505.00853.

[91] S. G. K. Patro and K.K. Sahu, "Normalization: A preprocessing stage," *arXiv preprint*, arXiv:1503.06462.

[92] C. Cheadle, M. P. Vawter, W. J.Freed, K. G. Becker, "Analysis of microarray data using Z score transformation," J. Mol. Diagn., vol. 5, no. 2, pp. 73-81, 2003.

[93] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*.

[94] G. Hinton, N. Srivastava, and K. Swersky, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited on*, vol.14, no. 8, 2012.

[95] D. P. Kingma, and J. Ba, "Adam: A method for stochastic optimization*," arXiv preprint*, arXiv:1412.6980.

[96] E. Brochu, V. M. Cora, and N. De Freitas, "A tutorial on bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint*, arXiv:1012.2599.

[97] C. E. Rasmussen, "Gaussian processes in machine learning," *In Summer School on Machine Learning (pp. 63-71). Springer*, 2003, Berlin, Heidelberg.

[98] L. Yang, and A., "On hyperparameter optimization of machine learning algorithms: Theory and practice*," arXiv preprint,* arXiv:2007.15745.

[99] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proc. IEEE*, vol. 104, no. 1, pp. 148–175, Jan. 2016, doi: 10.1109/JPROC.2015.2494218.

[100] W. Lyu et al., "An Efficient Bayesian Optimization Approach for Automated Optimization of Analog Circuits," in *IEEE Trans Circuits Syst. I*, vol. 65, no. 6, pp. 1954-1967, June 2018, doi: 10.1109/TCSI.2017.2768826.

[101] E. Brochu, V. M. Cora, and N. de Freitas, "A tutorial on bayesian optimization of expensive cost functions with application to active user modeling and hierarchical reinforcement learning," *arXiv preprint*, ArXiv abs/1012.2599.

[102] A. O'Hagan, "On curve fitting and optimal design for regression," *J. R. Stat. Soc B*, vol. 40, no. 1, pp. 1-42, 1978.

[103] J. Mocku, "Application of bayesian approach to numerical methods of global and stochastic optimization," *J Glob Optim*, vol. 4, no. 4, pp. 347-365, 1994.

[104] C. M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: *Springer*, 2007.

[105] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions*," J. Global Optim.,* vol. 13, no. 4, pp. 455–492, 1998.

[106] J. Mockus, V. Tiesis, and A. Zilinskas, "The application of bayesian methods for seeking the extremum," *Towards global optimization*, vol. 2, pp. 117-129, 1978.

[107] J. D. Rodriguez, A. Perez and J. A. Lozano, "Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 3, pp. 569-575, March 2010, doi: 10.1109/TPAMI.2009.187.

[108] Y. Jung, "Multiple predicting K-fold cross-validation for model selection," *Journal of Nonparametric Statistics*, vol. 30, no.1, pp. 197-215, 2018, doi:10.1080/10485252.2017.1404598.

[109] Y. Bengio, and G. Yves, "No unbiased estimator of the variance of k-fold cross-validation," *Journal of machine learning research* vol. 5, pp. 1089-1105, 2004.

[110] N. Calik, M. A. Belen, P. Mahouti and S. Koziel, "Accurate Modeling of Frequency Selective Surfaces Using Fully-Connected Regression Model With Automated Architecture Determination and Parameter Selection Based on Bayesian Optimization," *in IEEE Access*, vol. 9, pp. 38396-38410, 2021, doi: 10.1109/ACCESS.2021.3063523.

[111] N. Srivastava, G. Hinton, A. Krizhevsky,I. Sutskever and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929-1958, 2014.

[112] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *International conference on machine learning*, PMLR, vol. 37, pp. 448-456, 2015.

[113] Y. C. Chen, S. Y. Chen, and P. Hsu, "Dual-band slot dipole antenna fed by a coplanar waveguide," *IEEE Int. Symp. Ant. Prop.*, pp. 3589-3592, 2006, doi: 10.1109/APS.2006.1711396.

[114] S. Koziel and A. Bekasiewicz, "On reduced-cost design-oriented constrained surrogate modeling of antenna structures," *IEEE Ant. Wireless Prop. Lett.*, vol. 16, pp. 1618-1621, 2017, doi: 10.1109/LAWP.2017.2654304.

[115] Z. Hua, G. Haichuan, L. Hongmei, L. Beijia, L. Guanjun, and W. Qun, "A novel high-gain quasi-Yagi antenna with a parabolic reflector," *Int. Symp. Ant. Prop. (ISAP)*, Hobart, Australia, 2015.

[116] P. Mahouti, "Application of artificial intelligence algorithms on modeling of reflection phase characteristics of a nonuniform reflectarray element," *Int J Numer Model*, vol. 33, no. 2, 2020, doi: 10.1002/jnm.2689.

[117] Y. Zhang, X. Xu, "Solubility predictions through LSBoost for supercritical carbon dioxide in ionic liquids," *New Journal of Chemistry*, vol. 44, no. 47, pp. 20544-20567, 2020.

Slawomir Koziel received the M.Sc. and Ph.D. degrees in electronic engineering from Gdansk University of Technology, Poland, in 1995 and 2000, respectively. He also received the M.Sc. degrees in theoretical physics and in mathematics, in 2000 and 2002, respectively, as well as the PhD in mathematics in 2003, from the University of Gdansk, Poland. He is currently a Professor with the Department of Engineering, Reykjavik University, Iceland. His research interests include CAD and modeling of microwave and antenna structures, simulation-driven design, surrogate-based optimization, space mapping, circuit theory, analog signal processing, evolutionary computation and numerical analysis.

Nurullah Çalık received his M. Sc. And Ph.D. degree in Electronics and Communication Engineering from the Yıldız Technical University, Turkey, in 2013 and 2019, respectivly. He was worked as post-doctoral researcher in Informatics Institute in Istanbul Technical University. He is currently an Assistant Professor with the Department of Biomedical Engineering, Istanbul Medeniyet University, Turkey. The main research areas are large-scale data analysis, signal and image processing and AI applications in engineering. His research interests include especially deep learning regression, optimization and surrogate modelling.

Peyman Mahouti received his M. Sc. And Ph.D. degree in Electronics and Communication engineering from the Yıldız Technical University, Turkey, in 2013 and 2016, respectively. He is currently an Associated Professor with the Department of Electronic and Communication, Istanbul University - Cerrahpasa, Turkey. The main research areas are analytical and numerical modelling of microwave devices, optimization techniques for microwave stages, and application of artificial intelligence-based algorithms. His research interests include analytical and numerical modelling of microwave and antenna structures, surrogate-based optimization, and application of artificial intelligence algorithms.

Mehmet A. Belen received his Ph.D degree in Electronics and Communication engineering from the Yıldız Technical University in 2016. He is currently an Associated Prof. in İskenderun Technical University. His current activities include teaching and researching Electromagnetics and Microwaves along with developing Additive Manufacturing 3D Printed Microwave Components for Rapid Prototyping. His current research interests are in the areas of multivariable network theory, device modeling, computer aided microwave circuit design, monolithic microwave integrated circuits, and antenna arrays, active/passive microwave components especially in the field of metamaterial- based antennas and microwave filters.