Hyperparameter Optimization of Long Short-Term Memory Based Forecasting DNN for Antenna Modeling through Stochastic Methods

Lida Kouhalvandi, *Member, IEEE*, and Ladislau Matekovits, *Senior Member, IEEE*.

***Abstract*—This letter presents an impressive optimization method for determining the optimal model hyperparameters of a deep neural network (DNN) targeted to model the characteristics of antennas. In this paper we propose an innovative approach of efficient yield analysis for modeling and sizing antennas. It is based on the long short-term memory (LSTM) DNN aiming to forecast the extended frequency responses, where various stochastic methods are applied for determining the optimal hyperparameters while training a DNN. Among the various methods, the one which models the antenna accurately in terms of input scattering parameter, gain, and radiation patterns is the winner. The proposed method is compact and addresses the problem of heavy reliance to the designer experience in determining the hyperparameters. Additionally, forecasting the future frequency responses of the antenna reduces the designer's effort substantially in measuring large frequency band; hence, measuring whole frequency band would not be needed. For validating the effectiveness of the proposed method, the fabricated two element antenna array is used for modeling where the results demonstrate that the Thompson sampling (TS) algorithm can determine optimal hyperparameters with minimum error in comparison with other reported stochastic methods leads to predict the future frequency band accurately.**

***Index Terms*—antenna, deep neural network (DNN), forecasting, long short-term memory (LSTM), optimal hyperparameter, stochastic methods.**

I. Introduction

The next generation mobile communication networks, i.e., sixth-generation (6G) systems, are expected to be world-widely developed. Among other components, such systems need substantially high performance antennas [1]. Analyzing the various performances and specifications of antennas requires accurate and effective modeling techniques. Recently, machine learning and deep neural network (DNN) prove the validity in modeling sophisticated circuits; hence, they become popular in nowadays design methodology [2]. Even though the DNN is a sufficient modeling method, it requires systematical steps to be trained and constructed. One of the important and not straightforward consideration is achieving the optimal model hyperparameters that are required to construct the layers of any DNN includes input, hidden, and output layers.

Manuscript submitted****

(Corresponding author: ladislau.matekovits@polito.it).

L. Kouhalvandi is with the Department of Electrical and Electronics Engineering, Dogus University, Istanbul, Turkey.

L. Matekovits is with the Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Torino, Italy also with the Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni, National Research Council of Italy, 10129 Turin, Italy and with Politehnica University Timişoara, 300006 Timişoara, Romania.

For determining the optimal hyperprameters various strategies, as trial-error, grid search, random search, Bayesian optimization, genetic algorithm (GA) and particle swarm optimization (PSO) have been presented as methods of optimization [3], [4]. Even if these methods are useful in specifying the hyperparameters, however, various swarm intelligence and stochastic methods must be considered in order to configure the suitable method in optimizing various model hyperparameters of any DNN. In the recently reported literature, what is lack importantly is the investigation of various stochastic methods and the accuracy response of each of them. Secondly, what is appreciated in complex designs as antennas, is training the DNN that can provide responses for a large bandwidth. This expectation would not be provided straightforwardly and is missed in the recently reported literature [5].

This letter presents the investigation over various stochastic methods for concluding the appropriate algorithm for determining the optimal hyperparameter of the DNN that is used for modeling the behavior of antennas. To the best of authors' knowledge, the proposed method is for the very first time reported in the literature. In particular it is: 1) presenting the suitable method for optimizing the hyperparameters in terms of neuron number (NN), layer number (LN), learning rate (LR), dropout rate (DR), and batch size (BS) for antenna designs; 2) training the regression DNN by using the long short term memory (LSTM) layers for performing in the considered frequency range, and 3) forecasting the extended frequency band responses in terms of S_{11}, gain (G), E-plane and H-plane radiation patterns (RPs).

The paper is organized as following: Sec. II, presents concisely the proposed optimization theory. Section III is devoted to provide the experimental results of the applied method. Finally, Sec. IV concludes this work.

II. Optimization Method in a Nutshell

The DNNs, multi-layer neural networks, have proved their well-performance in various radio frequency (RF) applications leading to better accuracy rather than other reported methods in [6]. The essential need of wireless 6G networks is to provide wide-band designs. Hence, the performance of antennas in the whole frequency band must be considered. Such a design is time and memory demanding. Reduction of these aspects is beneficial and in turn reduces time-to-market part of the design. Moreover, forecasting the behaviour of the antennas in frequency band out of the interested range must be executed also for enhancing the overall performance of 6G communication systems including security and privacy perspectives as well [7].

For these reasons, we train and model the practical antenna with the regression DNN which includes the LSTM layers for considering the antenna's specifications in a large frequency band. The purpose of the LSTM layers is to accurately predict the EM responses beyond the frequency band of input sampling data. They are employed for learning long-term dependencies between frequency steps of sequences of data [8] includes frequency-dependent input layer features (i.e., S_{11}, gain, with E-plane and H-plane RPs). Additionally, forecasting DNN is constructed for predicting the future frequency series of antennas to be used in the 6G systems more trustfully. The significant question for any designer is this: how hyperparameters can be determined? Providing the optimal hyperparameters of any DNN is not straightforward and needs additional efforts. In order to have successful neural network, appropriate hyperparameters including LN, LR, DR, BS, and activation functions must be determined for controlling the behavior of trained NN. For this case, we list the various stochastic methods and apply them for predicting the optimal hyperparameters where finally we conclude which of the methods can be suitable for training LSTM-based DNN in terms of accuracy.

Figure 1 presents the general structure of proposed DNN for modeling the practical antenna's performances in terms of S_{11}, gain, E-plane and H-plane RPs where future frequency band is forecasted. For training a successful DNN, suitable model hyperparameters are required. Recently, the GA has been presented for hyperparameter optimization in power amplifier designs [3] and we extend this optimization by considering various swarm intelligence and stochastic methods reported in [9], [10] in terms of NN, LN, LR, DR, and BS. The various examined stochastic methods are: particle swarm optimization (PSO), artificial bee colony (ABC), ant colony optimization (ACO), firefly algorithm (FA), grey wolf optimizer (GWO), whale optimization algorithm (WOA), harris hawks optimizer (HHO), and Thompson sampling (TS). These stochastic methods are initialized with a random set of solutions and they are improved until finding the best solution. Generally, these methods are called multi-solution due to the existence of multiple solutions.

Algorithm 1 (at the end of this section) presents the summary of the proposed optimization method. The initial and the most important step of training any DNN is to provide suitable amount of dataset including training ($\mathrm{X}_{\mathrm{Train}}$), validation ($\mathrm{X}_{\mathrm{Val}}$), testing data ($\mathrm{X}_{\mathrm{Test}}$), and corresponding desired outputs ($\mathrm{Y}_{\mathrm{Train}}$, $\mathrm{Y}_{\mathrm{Val}}$, and $\mathrm{Y}_{\mathrm{Test}}$), here with a ratio of 70%, 15%, and 15%, respectively (Step-1). This amount of data is prepared by arranging a co-simulation environment between Microwave Studio (Dassault Systèmes) and numerical analyzer (such as MATLAB) [11]. Some design parameters of the antenna are determined and by using the co-simulation environment in the Microwave Studio, the generated data is transferred into the MATLAB environment. In simple words, the Microwave Studio is working in the background and MATLAB tool is handling all the generated data. These parameters are swept

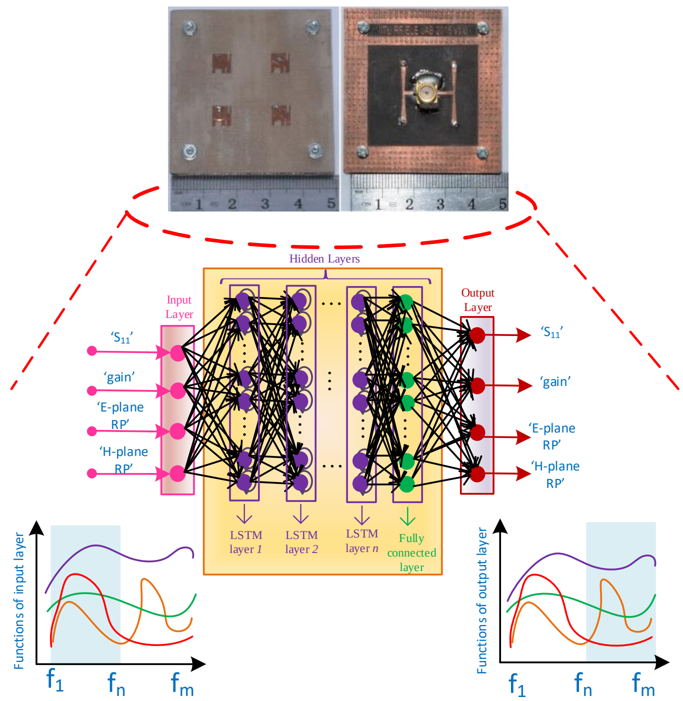

Figure 1: Proposed LSTM-based regression DNN for forecasting future frequency band used for modeling antennas.

within the range of [$\mp$5%-$\mp$25%] and the step size of $\mp$5%. The dataset is the combination of the values as $\mathrm{S}_{11_\mathrm{i}}$, G_i, E-plane RP_i and H-plane RP_i for $(i = 0, 1, ..., k)$ corresponding to the parameter setups of antenna, denoted as $\sum_i$. After generating dataset, input and output layer features are determined (Step-2) and then the LSTM-based hidden layers with fully connected layer are constructed for the regression DNN (Step-3). Afterwards, the hyperparameter specifications are defined and stochastic methods are employed regarding antenna's specifications as S_{11}, gain, and RPs. Each of these methods can be defined and constructed by getting knowledge from [12], [13] and using MATLAB tool [14] (Step-4). For our problem, the regression DNN can be trained and constructed using (1). The testing accuracy of this network can be achieved by calculating the difference between the actual testing outputs, i.e., Y_{Test}, and predicted outputs, i.e., Y_{Pred} (2) (Step-5).

$$\mathrm{net} = \mathrm{trainNetwork}(X_{\mathrm{Train}}, Y_{\mathrm{Train}}, \mathrm{layers}, \mathrm{options}) \quad (1)$$

$$Y_{\mathrm{Pred}} = \mathrm{predict}(\mathrm{net}, X_{\mathrm{Test}}) \quad (2)$$

Finally, the CPU execution environment is prepared and *'predictAndUpdateState'* function is employed [15] for forecasting the future frequency series of modeled antenna. The convergence of the proposed regression DNN is determined by the normalized root mean square error (RMSE) of testing data, presented as testing accuracy in the manuscript, is expected to be less that 1 (Step-6). In the proposed method, the rectified linear unit (ReLU) function is employed as the activation function, and the loss function is determined as the mean squared error for training the regression DNN.

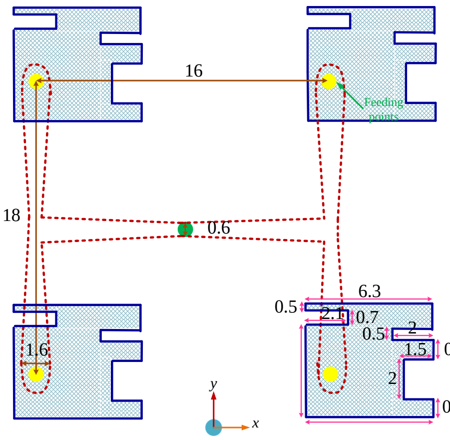

Figure 2: Designed 2×2 antenna array with the dimension sizes (width and length are in mm; drawing not in scale).

Algorithm 1 Overview of proposed method for determining optimal hyperparameters of LSTM-based DNN

1: Prepare X_{Train}, X_{Val}, and X_{Test} data;
2: Determine input and output layer features in a large frequency band (i.e., $[f_1...f_n...f_m]$) as S_{11}, gain, E-plane and H-plane RPs;
3: Construct LSTM layers following by the fully connected Layer;
4: Apply the stochastic methods for variables as: NN, LN, LR, DR, and BS;
5: Train the network using the *'trainNetwork'* option in MATLAB for forecasting the future frequency series;
6: If the RMSE value is < 1, then apply *'predictAndUpdateState'* for predicting the final and accurate frequency extended outcomes.

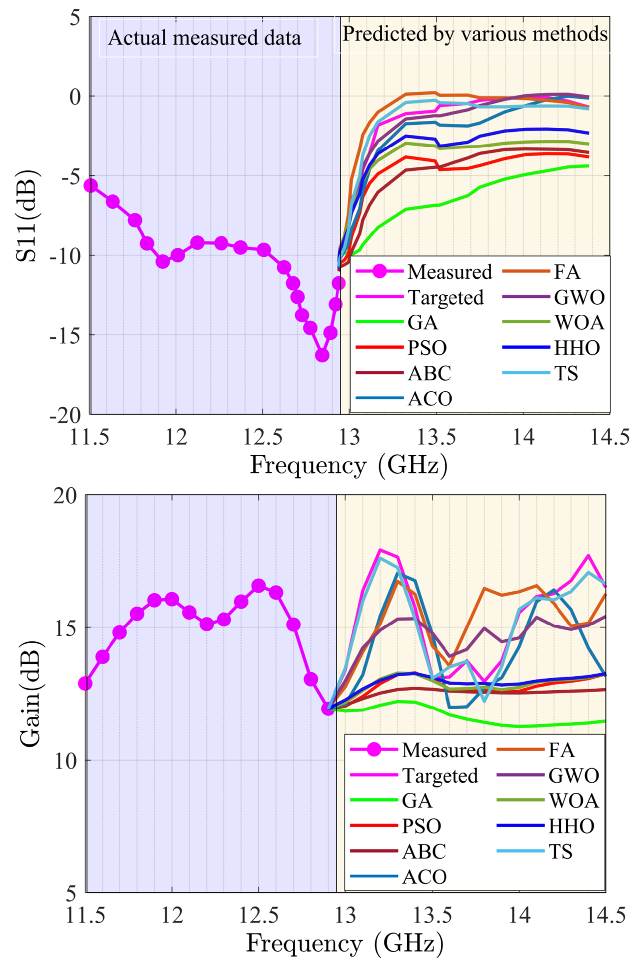

Figure 3: S_{11} and gain performances of fabricated antenna with measurement results (left side area and pink curve on the right hand side) and predicted performance by the proposed DNN (right side area).

III. Experimental Results

For validating the proposed method, we provide the measurement setup for the fabricated antenna array in Fig. 1 for various central frequencies where the detailed sizes are presented in Fig. 2. Additionally, the practical CPU execution environment includes an Intel Core i7-4790 CPU @ 3.60 GHz equipped with 32.0 GB RAM is used. The fabricated antenna is on the full ground plane where the four unit patch elements and array feed of antenna are realized using TSM-30-0600-C1/C2 (ε_r= 3) substrate with the height of 1.52 mm and 0.76 mm, respectively.

As the initial step, 2000 sequences include multi-segment S_{11}, gain, E-plane and H-plane RPs over the operation bandwidth are extracted from the arranged setup. Then, hidden layer structure as presented in Fig. 1 is provided. Afterwards, the training options are set as solver to *'adam'* and *'gradient threshold'* to 1. The Adam optimization algorithm and standard gradient descent algorithm are employed for updating the weights and biases of the network. Then, variables for optimization are set for achieving optimal NN, LN, LR, DR, and BS. These parameters are optimized using the nominated stochastic methods in Sec. II. These variables are optimized in the case of training a regression DNN leading to forecast the behaviour of the system in adjacent or far frequency band. The optimization methodology is generated and Tab. I presents the optimal achieved hyperparameters from each method. Figure 3 presents the S_{11} and gain performances in a large frequency band where from 11.5 to 13 GHz are the data achieved directly from the measurement; the range 13-14.5 GHz corresponds to the targeted output response in comparison with other predicted outcomes using various stochastic methods. This yellow region is predicted using the achieved hyperparameters in Tab. I, where the accuracy of regression DNN with the processing time of our proposed optimization process is presented in Tab. II. As it can be observed from this figure, the TS method can forecast the S_{11} performance more close to the targeted output response in the frequency band of 13-14.5 GHz.

For clarifying the effectiveness of the TS method, training accuracy and testing accuracy of the regression DNN with the optimal hyperparameters of 250 neurons in each '4' LSTM-layer are depicted in Fig. 4 where the testing accurcay is around 0.0439. Figure 5 summarizes the accuracy prediction

Table I: Antenna Modeling with the Proposed Methodology using Stochastic Methods

Method	**NN**	**LN**	**LR**	**DR**	**BS**
GA	70	1	0.02	0.3	0.8
PSO	110	2	0.01	0.5	1
ABC	100	1	0.01	0.5	1
ACO	180	3	0.007	0.6	0.95
FA	200	3	0.007	0.5	1
GWO	200	4	0.01	0.5	1
WOA	150	2	0.006	0.5	0.8
HHO	100	3	0.006	0.5	0.7
TS	250	4	0.005	0.5	1

Table II: Specifications of Trained DNN for Modeling Antenna through Stochastic Methods

Method	**Normalized testing accuracy**	**Processing time (min)**	**Normalized training accuracy**
GA	4.58	40	4.48
PSO	2.89	62	2.65
ABC	3.09	55	2.87
ACO	1.08	160	0.97
FA	1.02	188	0.84
GWO	0.74	202	0.67
WOA	1.96	105	1.78
HHO	1.55	125	1.48
TS	0.55	225	0.43

of various methods and illustrates that the TS method can predict the targeted output responses (i.e., S_{11}, gain, E-plane and H-plane RPs) more reliable than other methods with the overall normalized RMSE value of 0.55. The outcomes achieved from the TS method verifies the measurement results lead to suitable convergence to the E-plane and H-plane RPs. As Fig. 3 demonstrates, the large bandwidth is divided into two parts as 'actual measured data' and ' predicted by various method' in the two frequency bands of 11.5-13 GHz and 13-14.5 GHz, respectively. As it is results, the TS method can provide successful hyperparameters leads to forecast the half of the whole bandwidth, importantly.

In the paper results related to the input scattering parameter (matching) and gain are reported, since they are presenting a higher dynamics in their variation with respect to RPs. These later are less sensitive versus frequency, also considering the equal length feeding structure employed in the design (See Fig. 2). However, in case of other beam forming mechanisms, also allowing steering the beam, the variation in the RPs can be incorporated in the optimisation process with a higher weight.

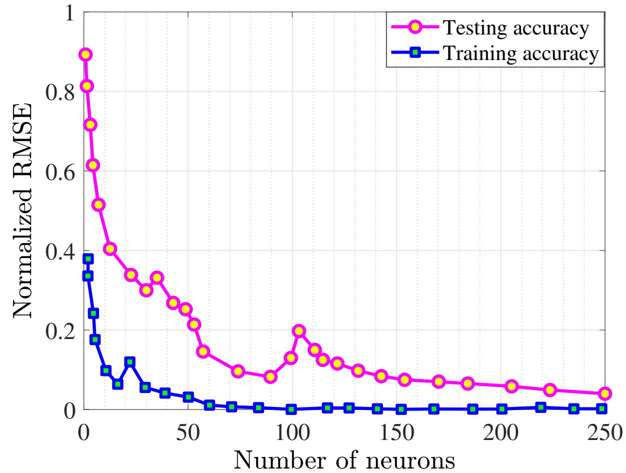

Figure 4: Normalized RMSE for training and testing data by employing the TS method for S_{11} performance.

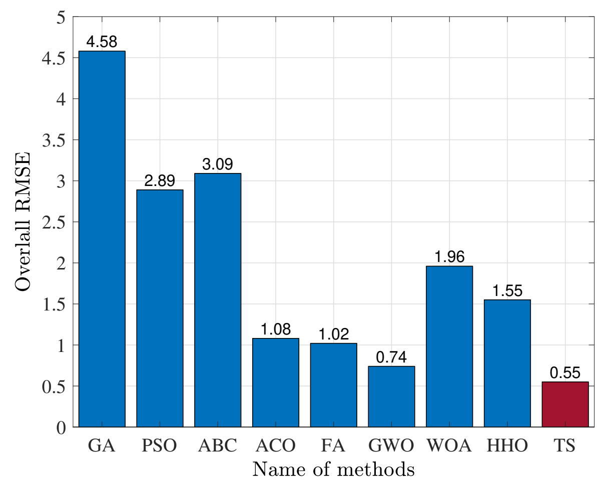

Figure 5: Overall normalized RMSE for providing testing accuracy achieved from each method.

IV. Conclusion

This letter presents two novel concepts in modeling the antennas through the DNN: determining the optimal hyper-parameters in constructing the LSTM-based DNN and predicting the extended output specifications of antennas. These contributions, proposed for the very first time in literature in the present form, help designers to construct the DNN reliable and to forecast the future antenna specifications lead to reduce effort in measuring, substantially. Various stochastic methods are employed for achieving the hyperparameters of the DNN that is for forecasting the future antenna performances in terms of S11, gain, and RPs. We train and construct the DNN on the fabricated 2×2 antenna array and the simulation results demonstrates that the TS method among the other reported methods is powerful enough to be applied in determine the hyperparameters of the LSTM-based DNN. Any antenna designer can employ this algorithm in antenna modeling for determining optimal hyperparameters lead to an accurate antenna's behavioral modeling and forecasting the extended antenna's specifications.

References

[1] C. D. Alwis, A. Kalla, Q.-V. Pham, P. Kumar, K. Dev, W.-J. Hwang, and M. Liyanage, "Survey on 6G frontiers: Trends, applications, requirements, technologies and future research," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 836–886, 2021.

[2] D. Gao, Q. Guo, and Y. C. Eldar, "Massive MIMO as an extreme learning machine," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 1046–1050, 2021.

[3] S. Wang, M. Roger, J. Sarrazin, and C. Lelandais-Perrault, "Hyperparameter optimization of two-hidden-layer neural networks for power amplifiers behavioral modeling using genetic algorithms," *IEEE Microwave and Wireless Components Letters*, vol. 29, no. 12, pp. 802–805, 2019.

[4] P. R. Lorenzo, J. Nalepa, L. S. Ramos, and J. R. Pastor, "Hyper-parameter selection in deep neural networks using parallel particle swarm optimization," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1864–1871. [Online]. Available: https://doi.org/10.1145/3067695.3084211

[5] F. Mir, L. Kouhalvandi, L. Matekovits, and E. O. Gunes, "Automated optimization for broadband flat-gain antenna designs with artificial neural network," *IET Microwaves, Antennas & Propagation*, vol. 15, no. 12, pp. 1537–1544, 2021. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/mia2.12137

[6] J. Jin, C. Zhang, F. Feng, W. Na, J. Ma, and Q.-J. Zhang, "Deep neural network technique for high-dimensional microwave modeling and applications to parameter extraction of microwave filters," *IEEE Transactions on Microwave Theory and Techniques*, vol. 67, no. 10, pp. 4140–4155, 2019.

[7] P. Porambage, G. Gür, D. P. M. Osorio, M. Liyanage, A. Gurtov, and M. Ylianttila, "The roadmap to 6G security and privacy," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 1094–1122, 2021.

[8] L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated deep neural learning-based optimization for high performance high power amplifier designs," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 67, no. 12, pp. 4420–4433, 2020.

[9] Q.-V. Pham, D. C. Nguyen, S. Mirjalili, D. T. Hoang, D. N. Nguyen, P. N. Pathirana, and W.-J. Hwang, "Swarm intelligence for next-generation wireless networks: Recent advances and applications," 2020.

[10] E. Bradford, A. M. Schweidtmann, and A. Lapkin, "Efficient multiobjective optimization employing gaussian processes, spectral sampling and a genetic algorithm," *Journal of Global Optimization*, vol. 71, no. 2, pp. 407–438, Jun 2018. [Online]. Available: https://doi.org/10.1007/s10898-018-0609-2

[11] L. Kouhalvandi, L. Matekovits, and I. Peter, "Key generation of biomedical implanted antennas through artificial neural networks," *In press at Proceedings of IEEE/ACM Conference on Connected Health Applications, Systems, and Engineering Technologies (IEEE/ACM CHASE 2021)*, 2021.

[12] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X19320941

[13] T. T. Ngo, A. Sadollah, and J. H. Kim, "A cooperative particle swarm optimizer with stochastic movements for computationally expensive numerical optimization problems," *Journal of Computational Science*, vol. 13, pp. 68–82, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877750316300047

[14] "Open source codes for generating optimization methods," https://nl.mathworks.com/matlabcentral/profile/authors/2943818, accessed: 2021-12-22.

[15] "Predict responses using a trained network," http://https://nl.mathworks.com/, accessed: 2021-07-16.