A Comparative Study of Surrogate Based Learning Methods in Solving Power Flow Problem

Oğuzhan Ceylan
Kadir Has University
Istanbul, Turkey
oguzhan.ceylan@khas.edu.tr

Gülşen Taşkın
Istanbul Technical University
Istanbul, Turkey
gulsentaskin@itu.edu.tr

Sumit Paudyal
Florida International University
Miami, USA
spaudyal@fiu.edu

***Abstract*—Due to increasing volume of measurements in smart grids, surrogate based learning approaches for modeling the power grids are becoming popular. This paper uses regression based models to find the unknown state variables on power systems. Generally, to determine these states, nonlinear systems of power flow equations are solved iteratively. This study considers that the power flow problem can be modeled as an data driven type of a model. Then, the state variables, i.e., voltage magnitudes and phase angles are obtained using machine learning based approaches, namely, Extreme Learning Machine (ELM), Gaussian Process Regression (GPR), and Support Vector Regression (SVR). Several simulations are performed on the IEEE 14 and 30-Bus test systems to validate surrogate based learning based models. Moreover, input data was modified with noise to simulate measurement errors. Numerical results showed that all three models can find state variables reasonably well even with measurement noise.**

***Index Terms*—power systems, machine learning, support vector regression, Gaussian process regression**

I. Introduction

With the increasing penetration of distributed generation sources and integration of renewables, the philosophy of the operation of the power systems is evolving. Development of the new communication technologies results in more and more measurements that brings big data in power systems. Motivated by these developments, machine learning methods, especially data driven models, have gained a lot of interest to replace complex nonlinear problems with simpler models, namely surrogate model or meta-model, derived from relationship of input and output samples of a physical system [1].

Various surrogate modeling methods have been introduced to effectively model the relations between the input and the output samples. Among them, Support Vector Regression (SVR) [2], Gaussian Process Regression (GPR) [3], and Extreme Learning Machines (ELM) [4] are the nonlinear multivariate methods providing promising and effective results in regression problems. The ELM [4] was introduced in 2006 to speed up conventional feed-forward neural networks by eliminating the time consuming iterative learning steps required for optimization of the input weights and output weights. Instead, the weights between the input and the hidden layer are randomly assigned in the Extreme Learning Machine, and the output weights are then directly calculated with a least squares based on the inverse matrix of the hidden layer output matrix of the neural network. The SVR [2] was developed based on statistical learning theory, which maps the input samples into higher dimensional nonlinear feature space using a kernel function, where the nonlinear problem is converted to a linear problem, resulting in higher discriminative features. The GPR [3] was formulated in terms of a Bayesian estimation problem, where the parameters are assumed to be random variables which are jointly drawn from a Gaussian distribution. Both GPR and SVR particularly work well when limited number of training samples are available.

Machine learning applications for power system problems are becoming increasingly popular. In [5], a general overview of the available techniques applicable in power systems domain are discussed. One of the initial attempts was aiming to use decision tree method in solving voltage security problems [6]. Later, machine learning methods were also applied to voltage stability prediction [7]. In [8] the authors have applied data-mining to detect the substations that are most sensitive to the disturbances. A neural network based methodology aiming to detect and classify power quality disturbances was developed in [9]. The authors of [10] predicted potential outage of power system components during hurricanes by using logistic regression method. Duration of power outages during hurricanes were predicted and compared in [11] by using several methods of regression models, and data mining techniques. Extreme learning machines was used for forecasting the outputs of photovoltaics in [12] and forecasting the outputs of wind turbines in [13]. In [14], the authors have used artificial neural networks and support vector machines to find the locations of faults in distribution systems. Similarly, machine learning approaches is used in in [15] solving dynamic security assessment problem.

This study approaches power flow problem as a black box model where the inputs are the active and reactive power injections/absorptions to/from the buses and the outputs are the voltage magnitudes and the phase angles of the buses. Based on available measurement data, surrogate based models are then generated by using three advanced regression methods, namely, Support Vector Regression (SVR), Gaussian Process Regression (GPR), and Extreme Learning Machines (ELM). The performances of the these methods for the IEEE 14- and 30-bus test Systems are evaluated based using R square and relative error.

II. Power Flow Problem

In power systems operation, power flow computations are vital to find the voltage magnitudes, bus angles, power flows, and power losses on the lines. The power systems operators need to perform power flow computations regularly to detect any possible problems in case of outages and take remedial actions on time. Power flow calculations are generally performed by using Newton Raphson method, where the system is modeled using nonlinear equations: $y = f(x)$. Note that, x represents the unknowns of the system state: phase angles (δ) and voltage magnitudes (V), and y represents the active and reactive power mismatch equations of the buses. Assuming a system with N buses, the active and reactive power equations for the k^{th} can be written following.

$$P_k(x) = V_k \sum_{n=1}^{N} Y_{kn} V_n cos(\delta_k - \delta n - \theta_{kn}) \tag{1}$$

$$Q_k(x) = V_k \sum_{n=1}^{N} Y_{kn} V_n sin(\delta_k + \delta n - \theta_{kn}) \tag{2}$$

In equations (1), and (2), V_k, Y_{kn}, θ_{kn}, and δ_k represents voltage magnitude of bus k, bus admittance between buses k and n, angle of the admittance between buses k and n, phase angle of voltage on bus k, respectively. For more details on the formulation and solution of the problem one may refer to [16].

This paper considers the power flow problem as a data-driven machine learning problem; hence, the system parameters are not required. It is assumed that the system state is determined by the changes of the active and reactive power injection/absorption from the buses. For a specific active and reactive power set, the outputs: voltage magnitudes of the buses and the angles are determined. One may think this model simply as following.

$$[v_i, \delta_i] = f(P_i, Q_i) \tag{3}$$

where v_i and δ_i represent the voltage magnitudes and the phase angles of the buses. P_i and Q_i represent the active and reactive power injections/absorptions.

III. Surrogate Modelling Methods

Given a set of data $X = \{(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)\}$ where $\mathbf{x_i} \in R^n$ is an input vector while y_i refers to associated continuous output derived from a physical system or a mathematical model. A surrogate model provides a function $f : x \to y$ to represent the relationship between the input and the output data as following.

$$y_i = f(X) + \epsilon_i \tag{4}$$

where ϵ_i is residue term and f is an unknown function determined by satisfying some optimization criteria. Support vector regression (SVR), Gaussian Process Regression (GPR), and Extreme Learning Machines (ELM) are some of the advanced regression methods using different loss functions to minimize, which we consider to model the power flow problem.

A. Support Vector Regression

Support Vector Regression is based on empirical risk minimization under Vapnik's $\epsilon-$insensitive loss function, which is defined as,

$$\mathcal{L}_\epsilon(y_i, f(x_i, w)) = \begin{cases} 0 & \text{if } |y_i - f(x_i, w)| \le \epsilon \\ |y_i - f(x_i, w)| - \epsilon & \text{otherwise.} \end{cases}$$

where ϵ is a parameter of loss function. The predictive model is chosen as,

$$f(x, w) = w^T \Phi(x) + b \tag{5}$$

In order to estimate the model parameters, w and b, the empirical risk minimization under the $\epsilon-$ sensitive loss function needs to be solved as,

$$\sum_{i=1}^{n} max(|y_i - w^T \Phi(x_i) - b| - \epsilon, 0) \tag{6}$$

This unconstrained optimization problem can be converted to a constrained optimization problem by including a regularization term in the objective function, yielding (7), so that it can be solved by its dual form to provide kernel calculations.

$$\begin{aligned} \underset{w,b,\xi}{\text{minimize}} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{n} (\xi_i, \xi_i')) \\ \text{subject to} \quad & y_i - w^T \Phi(x_i) - b \le \epsilon + \xi_i \\ & w^T \Phi(x_i) + b - y_i \le \epsilon + \xi_i' \\ & \xi_i > 0, \xi_i' > 0, i = 1. \end{aligned} \tag{7}$$

By following the Lagrangian, dual optimization problem is obtained and solved with Karush-Kuhn-Tucker conditions. The dual representation of the predictive model becomes,

$$f(x) = \sum_{i=1}^{n} (\alpha_i, \alpha_i^*) K(x, x_i) + b \tag{8}$$

where $\alpha_i, \alpha_i^* \in [0, C]$ coefficients refer to support vectors having non-zero coefficients. $K(x, x_i)$ is the kernel matrix constructed by radial basis function in this paper. It should be noted that this regression becomes radial basis network when using RBF kernel function, and the number of the support vectors equals to the number of hidden layers.

B. Gaussian Process Regression

Gaussian Process Regression, also known as Kriging method in statistics, is a probabilistic multivariate regression method, defined as a collection of random variables with a joint Gaussian distribution. For a given training set $D = \{(x_i, f_i), i = 1, \ldots, n\}$, where $f_i = f(x_i)$ refers to the value of the function at the location x_i, the GP is as follows,

$$f(\mathbf{x}) \sim GP(m(\mathbf{x_i}), k(\mathbf{x}_i, \mathbf{x}_j)) \tag{9}$$

where $\mathbf{x}$ and $k(\mathbf{x}_i, \mathbf{x}_j)$ correspond to mean and covariance function, respectively. The covariance matrix shows similarities between the pair of random variables, which can be built by means of squared exponential (SE) function,

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp(||\mathbf{x}_i - \mathbf{x}_j||^2 / (2\sigma^2)) \tag{10}$$

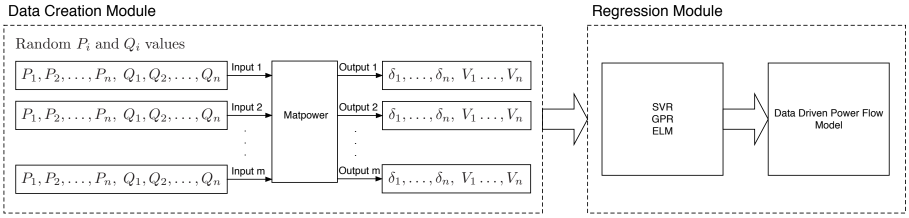

Fig. 1. Surrogate modelling schema for power flow problem.

where σ is the kernel width that needs to be tuned. The SE is the standard widely used covariance (kernel) function. Given an out-of-sample set $X*$, the desired predictive function f_* can be obtained as following,

$$p(\mathbf{f}_*|\mathbf{X}_*, \mathbf{X}, \mathbf{f}) = N(\mathbf{f}_*|\mu_*, \sigma_*) \quad (11)$$

where

$$\mu_* = \underline{k}_*^T.[K(X,X) + \sigma_n^2 I]^{-1}\underline{y} \quad (12)$$

$$\sigma_*^2 = k(\underline{X}_*, \underline{X}_*) - \underline{k}_*^T[K(X,X) + \sigma_n^2]^{-1}\underline{k}_*. \quad (13)$$

$\underline{k}_*$ is a vector of covariance values of training samples X and the out-of-sample $\underline{X}_*$, and σ_n^2 is the noise variance. σ_x^2 expresses the confidence measure associated by the model to output.

C. Extreme Learning Machines (ELM)

The predictive function in ELM for an input $\mathbf{x}$ is defined as following,

$$f_L(\mathbf{x}) = \sum_{i=1}^{L} w_i h_i(\mathbf{x}) \quad (14)$$

where L is the number of nodes at the hidden layer, and w_i is the output weight connecting the i-th hidden node to the output nodes. h_i represents a nonlinear feature mapping, which is defined as following,

$$h_i(\mathbf{x}) = G(\mathbf{a_i}, \mathbf{b_i}, \mathbf{x}) \quad (15)$$

where G is generally selected as a sigmoid function as.

$$G(\mathbf{a_i}, \mathbf{b_i}, \mathbf{x}) = 1/(1 + \exp(\mathbf{a}_i\mathbf{x} + b_i)) \quad (16)$$

The parameters a_i and b_i in (16) are the parameters of the nodes at the hidden layer, that are randomly generated in ELM to learn the weights w_i. Equation (14) can be rearranged in the matrix notation as following,

$$f_L(\mathbf{x}) = \mathbf{h}(\mathbf{x})^T\mathbf{w} \quad (17)$$

where $\mathbf{w} = [w_1, \cdots, w_L]$ and $\mathbf{h}(\mathbf{x})^T = [h_1(\mathbf{x}), \cdots, h_L(\mathbf{x})]$. Equation (17)

can be written in compact form as,

$$\mathbf{Hw} = \mathbf{Y} \quad (18)$$

where $\mathbf{Y}^T = [\mathbf{y_1}, \cdots, \mathbf{y_n}]$ is output vector of training set, and $\mathbf{H}$ is the hidden layer output matrix of the neural network expressed as

$$\mathbf{H} = \begin{bmatrix} h_1(\mathbf{x_1}) & \ldots & h_L(\mathbf{x_1}) \\ \vdots & \ddots & \vdots \\ h_1(\mathbf{x_n}) & \ldots & h_L(\mathbf{x_n}) \end{bmatrix}. \quad (19)$$

The optimal output weight, $\mathbf{w}$, is obtained by minimizing $||\mathbf{Hw} - \mathbf{Y}||^2$ in the least squares sense, yielding $\mathbf{w}^* = (\mathbf{H^T H})^{-1}\mathbf{H^T Y}$.

IV. Numerical Results

The surrogate modeling schema, consists of two parts, is shown in Fig. 1. As can be seen, in data creation module, initial step randomly creates different active and reactive loads to represent different operating states. In practical scenarios, these would actually come from measurement devices. Then, all those different states are solved by using Newton Raphson based load flow method in Matpower [17], and the solutions of the unknowns for each case which are the voltage magnitudes and the angles are retained. In the regression module, the input samples with their corresponding outputs are then fed to the regression algorithms, including GPR, SVR, and ELM, resulting in three surrogate models associated with each regression algorithm which represent power flow problem.

We used Matpower [17] to create data to simulate power flows. We used IEEE 14, and 30-bus test systems for the simulations. Two different type of tests for each test system were performed:

- First type of tests assume perfect information. For each test system 1,000 power flow cases with randomly created active and reactive powers are run and numerical results comprised of voltage magnitudes and angles are saved.
- Second type of tests included simulated noise in the input data.

Note that, for both IEEE 14 and IEEE 30-Bus systems, the input variables consisting of active and reactive power quantities are equal to two times the number of buses except the slack bus in the system. The number of outputs are determined by summing up the number of phase angles and number of load buses. Then, we implemented the regression methods. During the data creation process, maximum and minimum active and reactive load values were multiplied by constant factors, and those values were set as the ranges for the loads in all buses of the system. Also with inclusion of Gaussian White noise possible measurement errors were simulated in the input data.

In order to evaluate the performance of the surrogate models, the data created in data creation module were randomly divided into two parts by using two different training data:

- 20% of the training and 80% of the test data
- 50% percent of training and 50% of the test data.

All the unknowns of system, including voltage magnitudes and angles, are then estimated by associated surrogate models, and

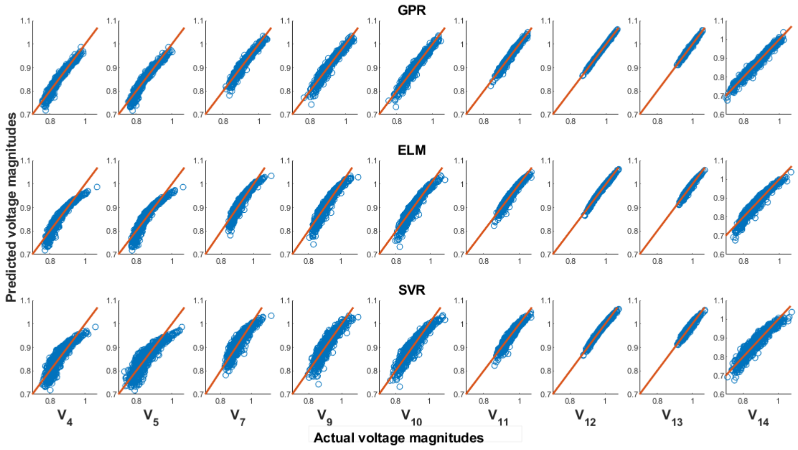

Fig. 2. Sample correlation plots of IEEE 14-Bus System variables for each surrogate model for voltage magnitudes with no noise training data=20%.

the goodness of fit for each model is analyzed with R^2 and relative errors in percentage on the test data.

Tables I and II show the relative % error of each surrogate model for voltage magnitudes and angles in the IEEE 14-Bus using training data of 20% and 50%, respectively. Similarly, Table III and Table IV represent the results for the IEEE 30-Bus system. Two sets of results are provided, the first one assumes that there is no noise in the input data, the second one assumes a signal to noise ratio of 5 on the measurement data. Results show that the measurement errors due to noise has an impact on the estimation errors; however, the three regression methods were able to find system parameters fairly accurately.

TABLE I
% ERROR OF EACH SURROGATE MODEL FOR VOLTAGE MAGNITUDES AND ANGLES ON THE IEEE 14-BUS SYSTEM, TRAINING DATA=20%.

No noise in data	**GPR**	**ELM**	**SVR**
Voltage Magnitudes	1.016	1.465	1.701
Angles	3.457	4.698	4.602
S/N ratio 5			
Voltage Magnitudes	1.103	1.506	1.635
Angles	3.826	4.894	5.118

TABLE II
% ERROR OF EACH SURROGATE MODEL FOR VOLTAGE MAGNITUDES AND ANGLES ON THE IEEE 14-BUS SYSTEM, TRAINING DATA=50%.

No noise in data	**GPR**	**ELM**	**SVR**
Voltage Magnitudes	0.740	1.357	1.454
Angles	2.637	4.452	4.412
S/N ratio 5			
Voltage Magnitudes	0.788	1.352	1.478
Angles	2.746	4.362	4.373

TABLE III
% ERROR OF EACH SURROGATE MODEL FOR VOLTAGE MAGNITUDES AND ANGLES ON THE IEEE 30-BUS SYSTEM, TRAINING DATA=20%.

No noise in data	**GPR**	**ELM**	**SVR**
Voltage Magnitudes	1.274	2.781	1.352
Angles	5.728	7.599	5.279
S/N ratio 5			
Voltage Magnitudes	1.257	2.461	1.332
Angles	6.004	7.200	5.406

TABLE IV
% ERROR OF EACH SURROGATE MODEL FOR VOLTAGE MAGNITUDES AND ANGLES ON THE IEEE 30-BUS SYSTEM, TRAINING DATA=50%.

No noise in data	**GPR**	**ELM**	**SVR**
Voltage Magnitudes	0.996	1.042	1.093
Angles	4.483	4.596	4.595
S/N ratio 5			
Voltage Magnitudes	1.077	1.141	1.259
Angles	4.897	5.098	5.103

Figures 2 and 3 show the correlation plots of the actual and predicted values of the voltage magnitudes and angles for the IEEE 14-bus test system for each surrogate models. A linear behaviour means that the actual and the predicted values are close to each other; hence, as the linearity increases on those graphs and associated method is assumed to successfully predict actual power flow results. From Fig. 2 it is easily seen that the graph related to the voltage magnitudes of the GPR method show the best linear behaviour, and the worst linear behaviour is obtained by the graph related to SVR. Similar behaviour is obtained for the angles in IEEE 14 Bus Test System: GPR provides the most accurate results, while SVR provides the worst ones.

Figures 4 and 5 show the R^2 values of IEEE 14 Bus and 30 Bus Test System results considering two different training data for both with noisy and non-noisy data for three different regression methods respectively. Note that in Figures 4 and 5, noisy and non-noisy data are represented with dashed and solid lines, respectively. GPR provides the best R^2 value for IEEE 14 Bus System. R^2 of the noisy data are slightly worse than those of non-noisy data. The differences in the results decrease in IEEE 30 Bus Test System case. Overall, SVR provides the worst results. It is obvious that training data size has a higher impact for IEEE 30 Bus Test System.

V. CONCLUSION

This paper explores the utility of state-of-art surrogate based modelling methods, including GPR, ELM and SVR in solving power flow equations. The effects of the size of training samples and data noise to surrogate modelling are analyzed.

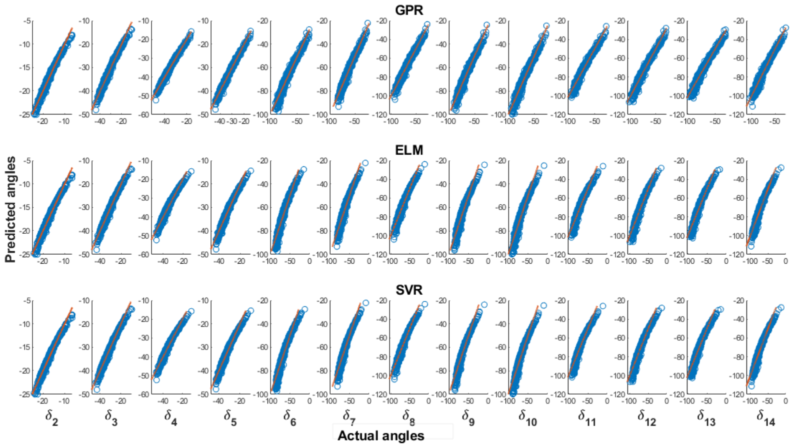

Fig. 3. Sample correlation plots of IEEE 14-Bus System variables for each surrogate model for angles with no noise training data=20%.

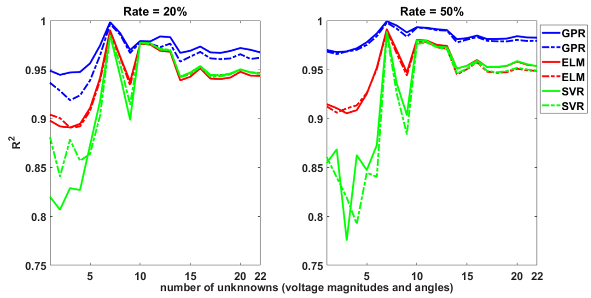

Fig. 4. 14 Bus Test System results: R^2 with different size of training data.

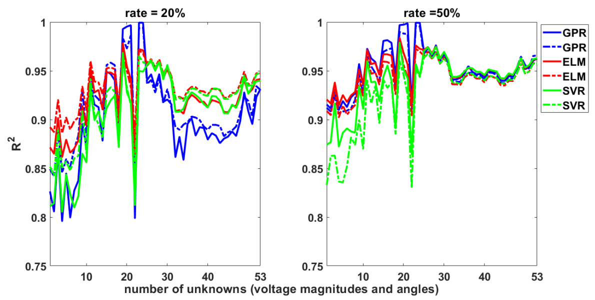

Fig. 5. 30 Bus Test System results: R^2 with different size of training data.

The results show that GPR provides more accurate and more robust predictions compared to the other regression methods. In order to make a deep analysis to reveal the effectiveness of surrogate modelling methods in power flow problem, more IEEE test systems will be included to the experiments along with including more evaluation criteria.

References

[1] R. E. Stern, J. Song, and D. B. Work, "Accelerated Monte Carlo system reliability analysis through machine-learning-based surrogate models of network connectivity," *Reliability Engineering and System Safety*, vol. 164, no. January, pp. 1–9, 2017.

[2] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep 1995.

[3] C. E. Rasmussen, "Evaluation of gaussian processes and other methods for non-linear regression," Ph.D. dissertation, Toronto, Ont., Canada, Canada, 1997.

[4] G.-B. Huang, Q.-y. Zhu, and C.-k. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[5] L. A. Wehenkel, *Automatic Learning Techniques in Power Systems*. USA: Kluwer Academic Publishers, 1998.

[6] T. Van Cutsem, L. Wehenkel, M. Pavella, B. Heilbronn, and M. Goubin, "Decision tree approaches to voltage security assessment," *IEE Proceedings C - Generation, Transmission and Distribution*, vol. 140, no. 3, pp. 189–198, May 1993.

[7] V. Malbasa, C. Zheng, P. Chen, T. Popovic, and M. Kezunovic, "Voltage stability prediction using active machine learning," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 3117–3124, Nov 2017.

[8] S. K. Tso, J. K. Lin, H. K. Ho, C. M. Mak, K. M. Yung, and Y. K. Ho, "Data mining for detection of sensitive buses and influential buses in a power system subjected to disturbances," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 563–568, Feb 2004.

[9] M. Valtierra-Rodriguez, R. de Jesus Romero-Troncoso, R. A. Osornio-Rios, and A. Garcia-Perez, "Detection and classification of single and combined power quality disturbances using neural networks," *IEEE Transactions on Industrial Electronics*, vol. 61, no. 5, pp. 2473–2482, May 2014.

[10] R. Eskandarpour and A. Khodaei, "Machine learning based power grid outage prediction in response to extreme events," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3315–3316, July 2017.

[11] R. Nateghi, S. D. Guikema, and S. M. Quiring, "Comparison and validation of statistical methods for predicting power outage durations in the event of hurricanes," *Risk Analysis*, vol. 31, no. 12, pp. 1897–1906.

[12] T. T. Teo, T. Logenthiran, and W. L. Woo, "Forecasting of photovoltaic power using extreme learning machine," in *2015 IEEE Innovative Smart Grid Technologies - Asia (ISGT ASIA)*, Nov 2015, pp. 1–6.

[13] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, "Probabilistic forecasting of wind power generation using extreme learning machine," *IEEE Transactions on Power Systems*, vol. 29, no. 3, pp. 1033–1044, May 2014.

[14] D. Thukaram, H. P. Khincha, and H. P. Vijaynarasimha, "Artificial neural network and support vector machine approach for locating faults in radial distribution systems," *IEEE Transactions on Power Delivery*, vol. 20, no. 2, pp. 710–721, April 2005.

[15] E. M. Voumvoulakis, A. E. Gavoyiannis, and N. D. Hatziargyriou, "Application of machine learning on power system dynamic security assessment," in *2007 International Conference on Intelligent Systems Applications to Power Systems*, Nov 2007, pp. 1–6.

[16] J. D. D. Glover and M. S. Sarma, *Power System Analysis and Design*, 3rd ed. Pacific Grove, CA, USA: Brooks/Cole Publishing Co., 2001.

[17] R. D. Zimmerman, C. E. Murillo-Sanchez, and R. J. Thomas, "Matpower: Steady-state operations, planning, and analysis tools for power systems research and education," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 12–19, Feb 2011.