21, rue d'Artois, F-75008 PARIS
http : //www.cigre.org

CIGRE US National Committee
2021 Grid of the Future Symposium

Photovoltaic Generation Estimation for DERMS Curtailment Analysis

M.G. HERNANDEZ, Y. LIU, A. VUKOJEVIC
Commonwealth Edison Company
USA

SUMMARY

Photovoltaic (PV) systems are widely used as an alternative energy generation to fossil energy. Excessive PV generation in one of the ComEd feeders may result in the backfeed on a substation transformer over the limit. ComEd deployed a Distributed Energy Resources Management System (DERMS) platform to avoid damage to the equipment by curtailing PV generation when the backfeed on the substation transformer is over the limit. It is important to estimate the amount of curtailed energy so that ComEd can evaluate the performance of DERMS. However, solar power generation is intermittent and stochastic as it is intrinsically highly dependent on weather fluctuations, making the calculation difficult. This study compares several techniques to estimate solar power generation using weather information from weather stations near PV plants installed in Chicago. Results show that Random Forest Regressor algorithm performs best in comparison with other techniques. In addition, we propose a method to utilize the estimated power to obtain the curtailed energy. Applying the methodology on May 1st, 2021 events shows the platform successfully curtailed PV energy within the required limit.

KEYWORDS

DER, curtailed energy, forecasting, photovoltaic, machine learning

melanygh9@gmail.com

I. INTRODUCTION

Renewable energy resources are the most promising alternative energy resources to reduce environmental pollution and harmful emissions caused by fossil energy [1]. Therefore, they have been widely integrated into power generation systems in the past decades. Solar energy is an easily accessible energy source that has proven to be one of the cleanest and most abundant renewable energy sources [2].

However, solar power generation is intermittent and stochastic because it depends on weather conditions [3]. The variable and hardly predictable nature of solar energy resources has increased solar energy penetration rates, which has generated difficulties for the management of electric power grids. These difficulties seriously affect the stable operation of the entire power system [2].

Distributed Energy Resource Management System (DERMS) is one of the emerging technologies to resolve the difficulties caused by DERs. As a pilot project, ComEd deploys a DERMS platform in a selected feeder where excessive PV generation may result in the backfeed on the substation transformer over the limit under certain conditions. The DERMS platform would curtail PV generation to reduce the backfeed when the backfeed is over the limit to avoid damage to the transformer. In the planning phase, a study using historical data showed that DERMS should curtail less than 5% energy annually.

To evaluate the performance of DERMS, a Metrics and Valuation process is being developed, which would require a good estimation of PV generation. The process utilizes the operation data and calculates the amount of energy curtailed during operation. The curtailed energy is calculated as the difference between the estimated generation without DERMS and the actual generation with curtailment triggered. Thus, PV power generation estimation plays a core role in the calculation to obtain an accurate amount of curtailed energy.

There are several PV power prediction methods that can be classified into two categories. The first category is classical or statistical methods, defined as data-driven methodologies to forecast the future behavior of PV cells by reducing the error and extracting features from historical samples [2, 3]. This category includes Autoregressive Integrated Moving Average (ARIMA) and Bass Diffusion. The second category refers to artificial intelligence techniques that have become an excellent tool for PV generation because they can solve the problem of a non-linear function estimation; they also avoid modeling complex atmospheric phenomena by focusing on the actual operation data.

In this study, power prediction techniques will be compared to choose the best approach. The selected methods are Linear Regression, Support Vector Regression, Long Short Term Memory Networks, and Random Forest Regressor. The performance is validated using real solar power collected in Chicago, Illinois. In addition, an evaluation of the curtailed energy by DERMS is performed to benchmark the performance of DERMS.

The paper is organized as follows. Section 2 describes the case of the study. Section 3 presents the different machine learning techniques that will be evaluated. In Section 4, the overall performance of the techniques is discussed as well as the analysis of curtailed energy in DERMS, followed by the conclusions in Section V.

II. CASE OF STUDY

As part of the deployment of DERMS at ComEd, an evaluation process has been developed to validate the performance of the platform. One of the most important evaluation goals is to estimate if the curtailed energy by DERMS is less than 5% amount of annual PV production established by the planning study.

The curtailed amount of energy is calculated as the difference between the estimated generation without DERMS and the actual generation with DERMS triggered during the curtailment event. Currently, the method to estimate generation without DERMS is to assume that the estimated generation during the curtailment event remains constant with the generation level at the last timestamp before DERMS triggered and curtailed the generation. This method, however, may result in an inaccurate overestimation or underestimation. Figure 1 shows how the energy curtailed (yellow area) could be underestimated with the current methodology.

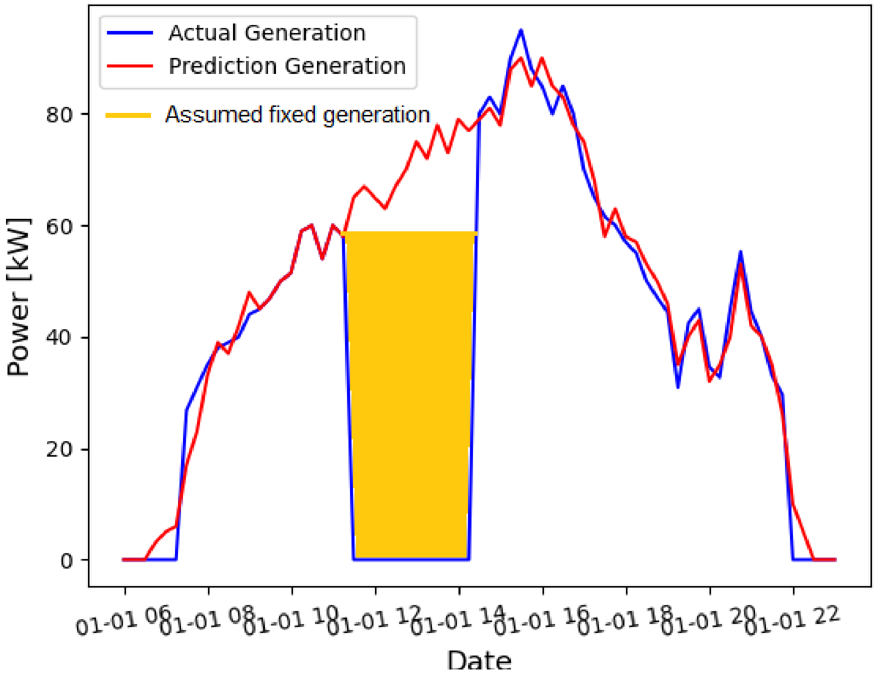

Figure 1: Illustrative Curtailment Data

To obtain a more accurate result, the following estimation process, shown in Figure 2, is developed in this study. First, the dataset is collected and preprocessed, then weather features and best performance estimator are selected in a data-driven way, and finally, the process calculates the curtailed energy percentage.

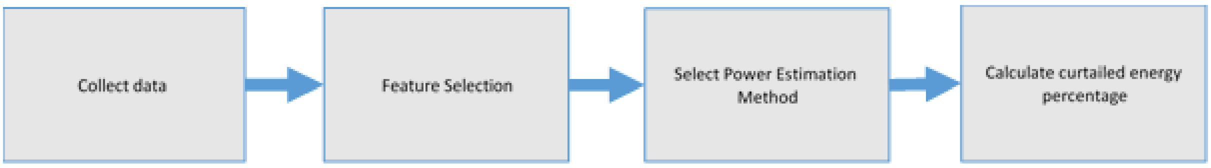

Figure 2: Flow Diagram of the proposed estimation process

A. *Dataset*

Two datasets were used in this study. First, A historical 15-minute-interval dataset of 2020 was utilized to select the most important features and the estimation algorithm. Then, weather data and PV Generation data from DERMS during April/May 2021 were used to train the estimation model selected using the first dataset and estimate the generation on May 1st, 2021, when the first curtailment event happened. Table I presents a list of the features in the datasets as well as their abbreviations. And Figure 3 shows some representation of the features from 2020 historical data.

Table I: List of Features and Abbreviations

Features	**Abbreviations**
Power [kW]	Pow
Global Horizontal Irradiance [W/m^2]	GHI
Direct Normal Irradiance [W/m^2]	DNI
Diffuse Horizontal Irradiance [W/m^2]	DHI
Plane of Array Irradiance [W/m^2]	PAI
Ambient Temperature [$°c$]	Temp
Wind Speed [m/s]	WSp
Liquid Precipitations	Prec
Solid Precipitations	SlPr
Time Stamp	Month, Day, Hour

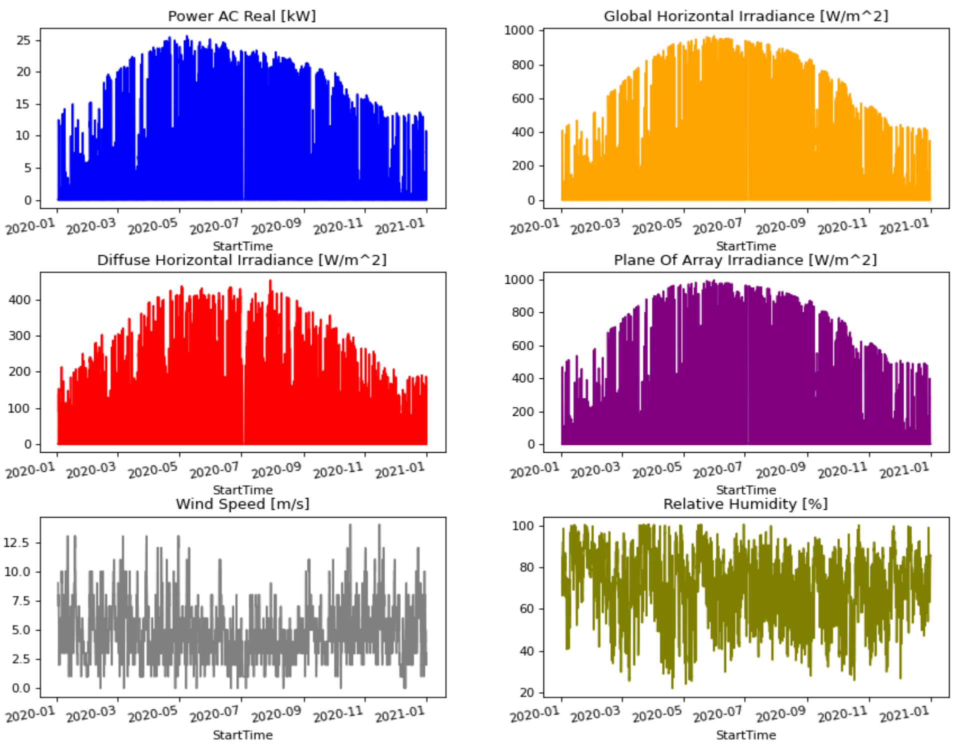

Figure 3: Representation of features from 2020 historical data

B. *Feature Selection*

Only the weather features most correlated with the generation are selected so that the process can minimize the number of features in the estimation. The heatmap shown in Figure 4 provides the correlation between power and the features, which is calculated as:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

where X, Y are random variables; μ_X, μ_Y are the expected values of X, Y respectively; and σ_X, σ_Y: standard deviation of X, Y respectively.

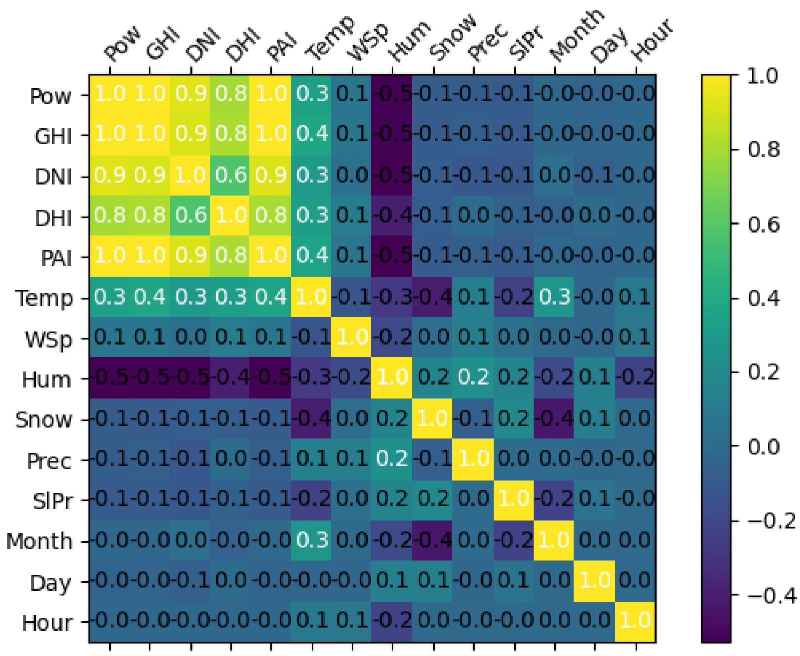

Figure 4: Correlation of features

From the figure, the following features show a strong correlation to the power generation: GHI, DHI, PAI, Temp, WSp, and Hum. Since Plane of Array Irradiance (PAI) is a linear combination of the rest of irradiances, utilizing PAI alone should include all the information from the rest of irradiances and can avoid overfitting while training the estimation algorithm. Although the features WSp, Snow, Prec, SlPr have the same correlation with the power in the 2020 historical data shown in Figure 4, a deeper study was developed for each month within the 2020 historical data, which reveals that WSp has a higher correlation with power than the other three features. Therefore, the selected features are PAI, Temperature, Wind Speed, and Humidity.

C. *Performance Metrics*

Several performance metrics were evaluated to select the best estimation method. The adopted metrics are mean absolute error (MAE), mean squared error (MSE), root mean square error (RMSE) defined as the squared root of MSE, and mean absolute percentage error (MAPE) [4]. If ŷ is the predicted value of the i-th sample, and y is the corresponding true value, then the different previous presented metrics estimated over N are defined as [5]:

$$MAE(y, \hat{y}) = \frac{1}{N}\sum_{i=0}^{N-1}|y_i - \hat{y}_i|$$

$$MSE(y, \hat{y}) = \frac{1}{N}\sum_{i=0}^{N-1}(y_i - \hat{y}_i)^2$$

$$MAPE(y, \hat{y}) = \frac{1}{N}\sum_{i=0}^{N-1}\frac{|y_i - \hat{y}_i|}{\max(\varepsilon, y_i)}$$

where ε is an arbitrary small yet strictly positive number to avoid undefined results when y is zero.

III. Estimation Techniques

In this paper, selected methodologies for solar power forecasting are evaluated. The focus of this section is on a case-specific description of each of the estimation techniques rather than their general theory representation.

A. Linear Regression (LR)

Linear regression is a simple but effective modeling technique when the real problem involves linear relationships. The main goal in LR is to identify the linear coefficients that relate the independent variables and a dependent variable [6, 7]. In this case, the independent variables are selected as the collected weather features, while the dependent variable considered is solar power. Linear Regression is computationally efficient; however, it is severely affected by outliers; therefore, it has low performance when the model is non-linear.

B. Support Vector Regression (SVR)

Support Vector Regressor adopts a Kernel function to map input patterns Xi into a higher dimension space. In this new space, output patterns become linearly separable, which allows extracting the pattern by linear fitting [7, 8]. SVR is memory efficient; however, it is sensitive to noise. The hyperparameters were tuned with the objective of finding the best fit. The penalty parameter C and Kernel function parameter γ were the most important parameters tuned as well as the Kernel used. To achieve the best performance, the selected C has the smallest value to avoid overfitting.

C. Long Short Term Memory (LSTM) Networks

Long Short Term Memory Networks is a special kind of Recurrent Neural Networks (RNN), capable of learning long-term dependencies. Among RNNs, Long Short-Term Memory networks have shown the best performance by learning information in the long term. Thus, in the past years, the interest in PV power prediction using LSTM networks has increased. RNN structure consists of the connection of each time step with the previous ones to incorporate the ability of prediction with sequence data. In the hidden layer, LSTM networks' structure has a special neuron called memory cell with the ability to store information over an arbitrary time differing from other RNN. LSTM networks perform well with the presence of irrelevant perturbations since they use a multiplicative input gate to control the memory units [9].

D. Random Forest Regressor (RFR)

Random Forest Regressor is considered an ensemble model since it integrates the regression outputs from several decision trees. A decision tree is a decision support tool that employs a tree-like model to estimate potential outputs for a given input. Implementation of an RFR is briefly described as follows: (i) Randomly generate T training sets S1, S2, …, ST from the original training dataset SN using the bootstrap method. (ii) Generate a decision tree C for each training set with added node split mechanism (iii) Employ the validation dataset X in each decision tree and produce multiple C1(X), C2(X), …, CT(X). (iv) The RFR result is obtained by averaging all the outputs from these trees [8]. In this study, hyperparameters such as the number of decision trees and amount of data in each subset were tuned, and the number associated with a minimum forecasting error was selected.

IV. Numerical Performance and Discussion

In this section, the numerical performance of the different used algorithms is discussed as well as the analysis of the curtailment Event of May 1st, 2021. The training dataset is normalized between 0 and 1 to eliminate scale differences. All algorithms are implemented in Python 3.9.5.

A. Selection of Algorithm

Historical 2020 data is split into train and test datasets, 75% and 25% of the data, respectively. Train data is applied to train the models, and the test dataset is applied to calculate the metrics in section II.C. Table II shows the error results obtained by each of the methods, as well as the time that each algorithm spent to compute the output. It is noticeable that Random Forest Regressor has the best performance among all the other methods. Thus, the selected method to estimate the power generation and calculate curtailed energy is RFR.

Table II: Algorithms Numerical Performance

	LR	SVR	LSTM	RFR
MAE	0.408	0.167	0.158	0.036
MSE	0.226	0.050	0.043	0.004
RMSE	0.476	0.223	0.208	0.065
MAPE	19.37	9.15	5.93	0.005
Time [s]	0.62	19.41	168.25	17.79

Figure 5 shows a representation of the estimated PV generation during November 2020, remarking the prediction interval [10]. The result shows that RFR gives a good estimation of the November data using the features selected.

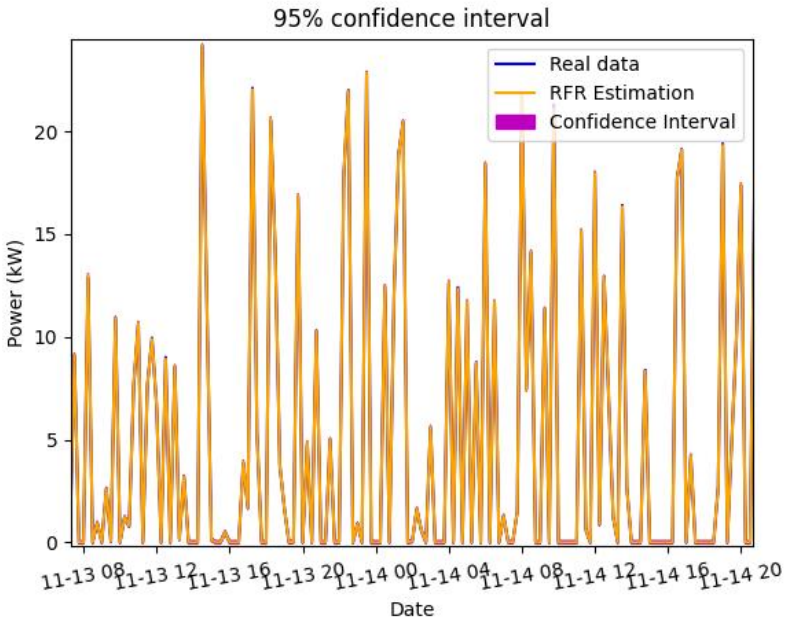

Figure 5: Illustration of the confidence interval for RFR

B. Analysis of curtailment event

The formula used to calculate the curtailment in event n at each time t is the difference of the estimated generation and the curtailed output:

$$Curt_{n,t} = Prediction_t - Actual_t$$

Total curtailment is defined as the sum of curtailment for all the time in the interval of event n and then the sum of the energy curtailed during all the events in the evaluation period:

$$E_{curt} = \sum_{\mathrm{n}=0}^{N} \sum_{t=0}^{T} Curt_{n,t} t$$

In this study, the percentage of the curtailed energy was calculated by:

$$\mathit{Curtailment\ Percentage} = \frac{\text{Total energy curtailed}}{\text{Total generation of that month}} * 100$$

On May 1st, 2021, the first curtailment event since the deployment of DERMS occurred. After training the RFR model with April and May data in 2021, the model is used to estimate solar generation on May 1st. The comparison of estimated generation and actual generation is visualized in Figure 6. The difference, which is the highlighted area, represents the curtailed energy during the event. Using fixed generation assumption, the estimated curtailment is 2.41 MWh. However, the result using the RFR model shows that the total energy curtailed during the event of May 1st, 2021 was 2.796 MWh, demonstrating that the curtailment was underestimated using fixed generation assumption. Since the total generation of that month was 279.720 MWh, the curtailment percentage of that month is calculated as 0.99% monthly.

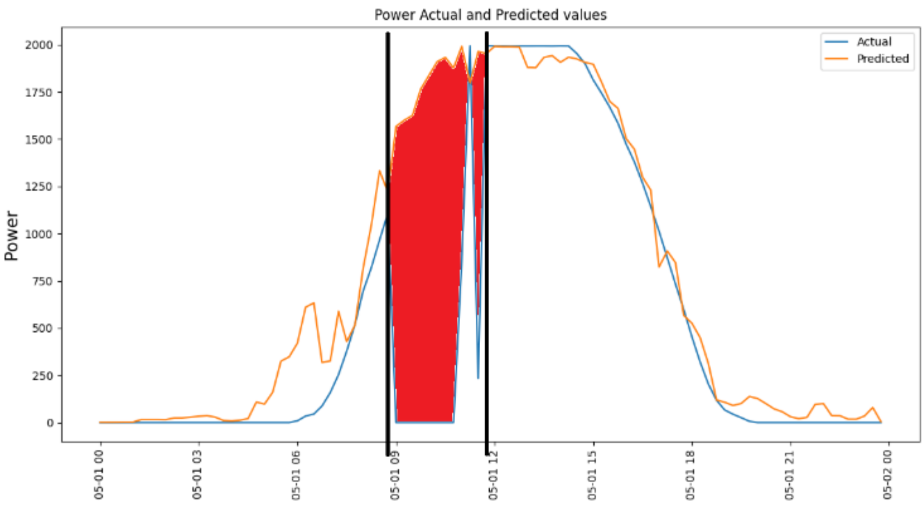

Figure 6 Curtailment calculation for May 1st, 2021

V. Conclusions

A new strategy to calculate the curtailed energy was proposed by using DER generation estimation techniques, which would be more accurate than the previous calculation strategy by fixed assumption. The features for estimation were selected based on correlation techniques in a data-driven way. The extensive results show that the Random Forest Regressor method can accurately predict the output power of PV, showing the best performance among the employed estimators. Analysis of curtailment events on May 1st reveals that the level of curtailment is below the limits. Therefore, the DERMS platform is working as expected. In future work, we propose to employ the algorithm with a complete year of data to be able to evaluate the behavior of DERMS over that period.

BIBLIOGRAPHY

[1] Wang, H., et al.: A review of deep learning for renewable energy forecasting. Energy Convers. Manag. 198 (2019)

[2] Gangqiang Li & Huaizhi Wang & Shengli Zhang & Jiantao Xin & Huichuan Liu, 2019. "Recurrent Neural Networks Based Photovoltaic Power Forecasting Approach," Energies, MDPI, Open Access Journal, vol. 12(13), pages 1-17, July.

[3] Seul-Gi Kim & Jae-Yoon Jung & Min Kyu Sim, 2019. "A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning," Sustainability, MDPI, Open Access Journal, vol. 11(5), pages 1-16, March.

[4] Runge, J.; Zmeureanu, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. Energies 2019, 12, 3254. https://doi.org/10.3390/en12173254

[5] De Giorgi, Maria Grazia, Congedo, Paolo Maria, Malvoni, Maria, & Laforgia, Domenico (2015). Error analysis of hybrid photovoltaic power forecasting models: A case study of Mediterranean climate. Energy Conversion and Management, 117-130. doi:101016/jenconman201504078

[6] Moon, Jihoon & Park, Sungwoo & Rho, Seungnim & Hwang, Eujun. (2019) A comparative analysis of neural Network architectures for building energy consumption forecasting. International Journal of Distributed Sensor Networks. vol. 15, September.

[7] S. Preda, et al., "PV Forecasting Using Support Vector Machine Learning in a Big Data Analytics Context," Symmetry, vol. 10, 2018.

[8] D. Su, E. Batzelis and B. Pal, "Machine Learning Algorithms in Forecasting of Photovoltaic Power Generation," 2019 International Conference on Smart Energy Systems and Technologies (SEST), 2019, pp. 1-6, doi: 10.1109/SEST.2019.8849106.

[9] Lara-Benítez, Pedro & Carranza-García, Manuel & Riquelme, José C. "An Experimental Review on Deep Learning Architectures for Time Series Forecasting", International Journal of Neural Systems, vol. 31(3), Feb 2021.

[10] A. Khosravi, S. Nahavandi, D. Creighton and A. F. Atiya, "Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances," in IEEE Transactions on Neural Networks, vol. 22, no. 9, pp. 1341-1356, Sept. 2011, doi: 10.1109/TNN.2011.2162110.