Short-Term Prediction of Solar Photovoltaic Power Generation Using a Digital Twin

John Yonce,
IEEE Student Member
Real-Time Power and Intelligent Systems Laboratory
Holcombe Department of Electrical and Computer Engineering
Clemson University,
Clemson, SC 29634, USA
johnyonce@ieee.org

Michael Walters,
IEEE Student Member
Real-Time Power and Intelligent Systems Laboratory
Holcombe Department of Electrical and Computer Engineering
Clemson University
Clemson, SC 29634, USA
mawalters@ieee.org

Ganesh K. Venayagamoorthy,
IEEE Fellow
Real-Time Power and Intelligent Systems Laboratory
Holcombe Department of Electrical and Computer Engineering
Clemson University
Clemson, SC 29634, USA
gkumar@ieee.org

***Abstract*— Large volumes of distributed energy resources (DERs), such as solar photovoltaic (PV) plants are integrated into the power distribution system due to increased awareness of climate change. These DERs introduce variable and uncertain generation sources due to changing weather conditions. This makes operations and controls challenging and complex. To better understand and manage the dynamic nature of solar PV power plants, digital twins (DTs) will be needed. DTs based on artificial intelligence (AI) methods can be applied to replicate the dynamics of PV plants. This study utilizes a popular paradigm of AI - neural networks to create a variety of data-driven DT (DD-DT) prediction models for a 1 MW solar PV plant located at Clemson University in South Carolina, USA. State-of-the-art internet of things (IoT) based real-time measurements are used to develop the DD-DTs. Typical results for short-term PV power prediction for DTs implemented using multilayer perceptron neural networks (MLPNNs) and Elman recurrent neural networks (ERNNs) are presented in this paper.**

***Index Terms*—Artificial intelligence, digital twin, neural networks, prediction, solar PV**

I. Introduction

Conventional electric power generation sources are being challenged by their negative global effects on climate, worldwide decarbonization legislation and efforts to modernize the power grid. These factors have enhanced the drive to integrate clean renewable energy sources (RESs). The development and increasing level of usage of RESs has further ushered in the inclusion of advanced distribution systems, such as distributed energy resources (DERs). Solar photovoltaic (PV) power production is at the forefront of DER development. However, challenges arise in power grid operations and controls due to the uncertainty and variability of weather conditions. Thus, full utilization of DER generation techniques is particularly difficult without the foresight of predictive modeling. Predictive modeling of PV plants in dynamic and uncertain environmental conditions can reduce the complexity in operations and management of the power system, improving efficiency and resiliency. Coupled with their widespread applications and the rapid growth of solar PV generation technology, predictive modeling has become a critical topic in research. Multi-timescale predictive modeling provides a variety of applications, including optimal energy dispatch, system health, monitoring, predictive asset maintenance, and planning and expansion of the power system [1].

Digital twins (DTs) have gained popularity across multiple disciplines due to their capability to link components from the physical and digital worlds based on physical properties or data. A data-driven DT is built on the basis of real time and historical data, providing an up-to-date, reliable reproduction of attributes and behaviors of physical systems within a virtual environment. The foundations of DTs were originally introduced in NASA's Apollo program for product life cycle testing when physical components were not available [2]. Since then, DTs have made a significant impact in several industries, especially in manufacturing by providing a platform to virtually represent factories, resources and workforces [2]. Due to the increasing complexity of power systems with DERs, DT applications are emerging, including for model validation, dispatch optimization, outage planning and forecasting [3, 4].

DT technology can be used for predictive applications such as solar power forecasting. The link between physical and virtual systems allows the plant model to adapt to changing conditions more quickly and accurately than a conventional physics-based model. In addition to adaptability, Data-driven DTs offer a high degree of system identification with fewer parameters [5].

In this study, a digital twin model of a 1 MW solar PV plant located at Clemson University, South Carolina, USA is developed utilizing neural networks, a popular paradigm of artificial intelligence (AI). The DT is implemented for short-term solar PV power predictions. State-of-the-art internet of things (IoT) based real-time measurements are used to develop

This research was funded by US NSF grant CNS 2131070 and the Duke Energy Distinguished Professor Endowment Fund. Any opinions, findings and conclusions or recommendations expressed in this material are those of author(s) and do not necessarily reflect the views of NSF and Duke Energy

the DTs. Typical results for multilayer perceptron neural networks (MLPNNs) and Elman recurrent neural networks (ERNNs) are presented. The rest of the paper is as follows: Section II describes DTs for solar PV power predictions. The implementation of DTs using neural networks is described in Section III. Section IV presents typical results and discussions. Finally, Section V provides the conclusion and future work.

II. Digital Twin For PV Power Prediction

Digital twins have gained momentum in both academic research and industrial applications due to their modeling capabilities. DTs are defined as having a physical component, virtual component and an interconnection between the two components. They may be built on the basis of physical principles and/or measured data [6]. The physical component provides a foundation for properties and data collection, on which the virtual replica is created and updated, with an emphasis on utilizing real-time data. Technology innovations in IoT and artificial intelligence have further synthesized digital and physical worlds by providing high-volume data collection and advanced computational performance capabilities, respectively [4]. Since DTs may be used to model and simulate a variety of physical processes and/or systems, they have been utilized in monitoring, control, predictive maintenance, risk management, and decision-support applications across multiple industries [7]. Specifically in power systems, digital twins offer a platform with the potential to revolutionize operations and control. System operators in energy control centers are challenged by the evolving generation technology landscape, especially with RESs and DERs. DTs provide a gateway to address operational challenges associated with these generation sources by establishing advanced planning and scheduling for energy dispatch, thereby increasing situational awareness in these centers [8].

Renewable energy sources have had an increasing share in today's energy market, providing distributed generation opportunities that fuel innovations in the smart grid. More specifically, various RESs may be combined to support smart cities containing solar PV-powered houses, smart buildings and electric vehicles [7]. No matter the size or technology used, each of these individual projects have specific properties relevant to both location-based weather conditions and specifications of generating sources; no two sites are the same [8]. Solar PV plants are a common RES utilized at these sites. However, the reliance on exterior environmental conditions presents a large barrier in large scale distributed implementation of these systems in power systems generation [8].

Digital twins are one of the most promising technologies to bridge the gap between large scale dynamic solar PV generation implementation and optimal decision-making in control centers, due to their predictive modeling capability. Data-driven DT (DD-DT) modeling is a promising approach, since data is a representation of both known and unknown physical parameters. The DD-DT model can thereby account for the full physical state of the PV plants, without prior knowledge of individual characteristics of solar PV plants [7]. Furthermore, the use of IoT and AI-based resources will further improve DD-DT models. IoT provides data in real-time, enhancing the quality of data. AI then utilizes this data to learn the behaviors and characteristics of the physical system within the virtual environment.

In Fig. 1, a model of the digital twin developed for Clemson University's 1 MW solar PV plant is shown. This model consists of three components as follows: Clemson University's R06 site parking lot, Real-Time Power and Intelligent Systems Laboratory (RTPIS Lab) and Clemson University's local area network (LAN), corresponding to the physical, virtual and communication components, respectively. The R06 parking lot contains the 1 MW solar PV plant, and IoT devices (micro-PMU and weather station), providing real-time data of the current physical state. These parameters are streamed via the LAN network to the RTPIS Lab, where the DT is implemented to carry out prediction utilizing AI algorithms.

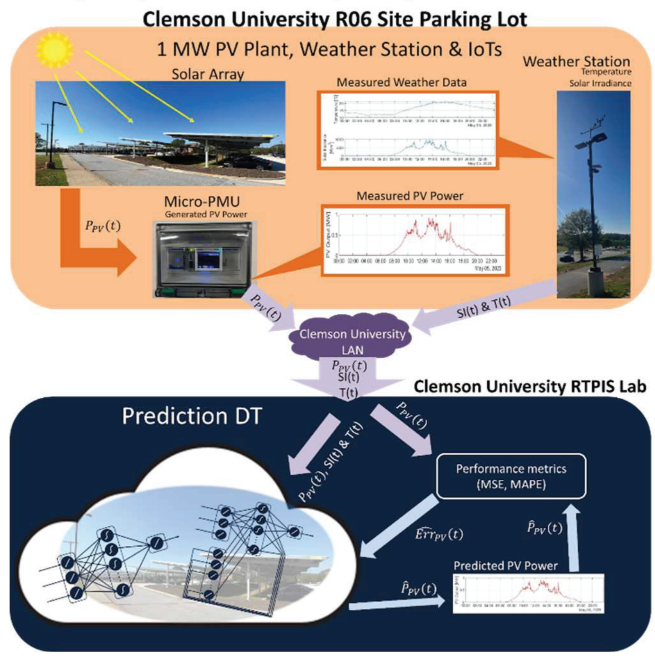

Figure 1. Model of Clemson University's R-06 Site and DT.

III. Implementation of DT-Prediction Using Neural Networks

Neural networks have proven to excel at PV power prediction [9]. For the virtual side of the digital twin, two neural network algorithms were utilized. The MLPNN architecture consists of a feedforward algorithm, depicted in Fig. 2. In contrast, the ERNN architecture features a recurrent algorithm. With ERNNs, outputs of the hidden layer from the previous timestep are used as additional inputs at the current timestep, adding a memory component to the network. A depiction of ERNN architecture is depicted in Fig. 3. These networks are presented with different input vectors, where the MLPNN input vector includes temperature (T), solar irradiance (SI) and generated PV power (P_{PV}) of the current and three previous timesteps. The ERNN only includes present temperature, solar irradiance, and generated PV power. In the ERNN, previous values from the decision vector are included in the input vector. For both networks, input and output layers consist of linear neurons and the hidden layer consists of 25 sigmoidal neurons.

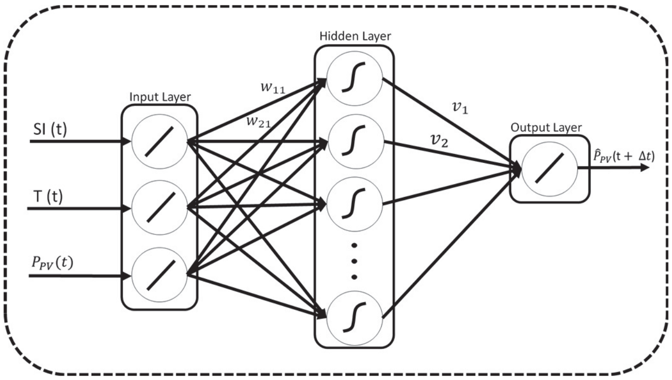

Figure 2. MLPNN architecture diagram. Outputting PV power prediction. No context layer or feedback.

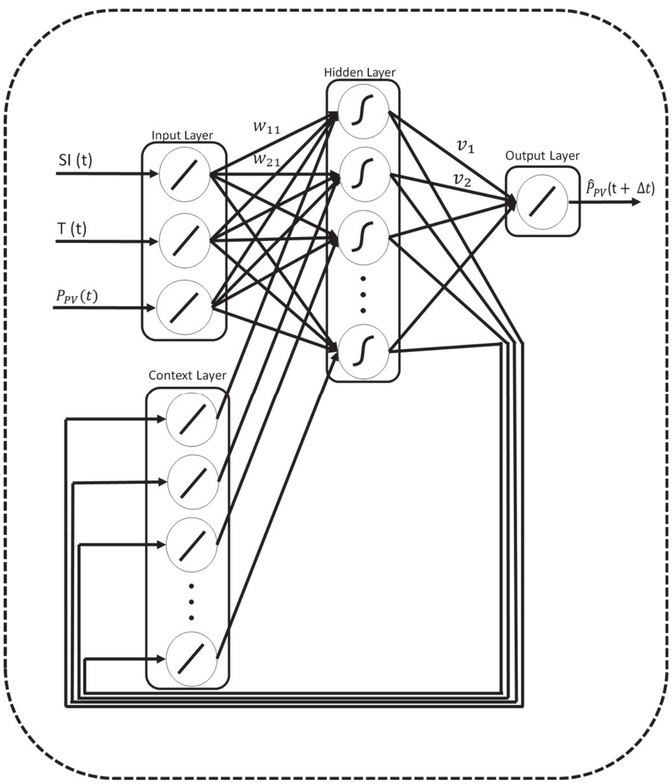

Figure 3. ERNN architecture diagram. Outputting PV power prediction. Context layer loops decision vector back into input.

Both neural networks are trained with a historical dataset consisting of 86 days ranging from March 2023 to June 2023. Days missing significant windows of data were discarded, but other methods such as data interpolation could be used to preserve data [10]. The dataset of spring to early summer provides a range of solar irradiance and temperature profiles. Different seasons offer different ranges of solar irradiance and temperature but ultimately different profile categories are utilized to better account for differences in weather patterns. Alternate sites with different weather conditions will require a separate digital twin model, but all required data can be obtained through a similar weather station. Input values are normalized to the range [0, 1]. The backpropagation algorithm discussed in [11] is used as the training method for both types of neural networks. The input vector (***x***) is assembled using *T*, *SI*, and P_{PV} at time t. Then, the input weights (***w***) are multiplied by the input vector to compute the activation matrix (***a***), as shown in (1).

$$a = w \times x(t) \tag{1}$$

This is directly followed by computing the decision matrix (***d***) using a sigmoid function given in (2).

$$d = \frac{1}{1+e^{-a}} \tag{2}$$

The next step is calculating the predicted PV power ($\hat{P}_{PV}$). This is done by multiplying the output weights (***v***) by the decision matrix as seen in (3).

$$\hat{P}_{PV}(t+\Delta t) = v \times d \tag{3}$$

Backpropagation uses the error (e_y) between actual PV plant power (P_{PV}) and its predicted values as shown in (4) to calculate the neural network weight changes. The trained weights are then used in forward propagation to compute the predictions.

$$e_y = P_{PV}(t) - \hat{P}_{PV}(t) \tag{4}$$

The prediction error is then used to calculate the activation error (e_a) and the decision error (e_d) as seen in (5) and (6), respectively.

$$e_d = v^T e_y \tag{5}$$

$$e_a = d(1-d)e_d \tag{6}$$

The activation error and decision error are then used to calculate input and output weight updates (Δw and Δv) using momentum (γ_m) and learning rate (γ_g) as seen in (7) and (8).

$$\mathrm{w}(t+\Delta t) = \mathrm{w}(t) + \gamma_m * \Delta\mathrm{w}(t-1) + \gamma_g * e_y * d^T \tag{7}$$

$$\mathrm{v}(t+\Delta t) = \mathrm{v}(t) + \gamma_m * \Delta\mathrm{v}(t-1) + \gamma_g * e_a * x^T \tag{8}$$

The weights are then used in forward propagation to compute the predictions.

To better account for the volatility of solar irradiance, each day in the historical training dataset is classified into four different categories based on its generated PV power: sunny, partly cloudy, moderately cloudy, and mostly cloudy. Since each category of day has a distinct measured PV profile, an optimized network with tuned parameters is developed and utilized. In addition to categorizing weather profiles of individual days, attention-based training is utilized for further parameter fine-tuning during the training stage. Attention-based training focuses on improving intervals of high error by amplifying the error signal by a constant greater than one. Therefore, the modified error signal is much larger over these intervals, increasing the impact during backpropagation.

IV. Results and Discussion

The testing data consists of the final seventeen days of the dataset. Four specific days were chosen based on their weather profiles. Classifications are as follows: May 25th and 26th as partly cloudy, May 27th as moderately cloudy and May 28th as mostly cloudy. Each of these days offers a variety of temperature, solar irradiance and measured generated PV power generation. Figs. 4 and 5 show DT-predictions utilizing MLPNNs and Figs. 6 and 7 show DT-predictions utilizing ERNNs, both with a 5-minute time horizon.

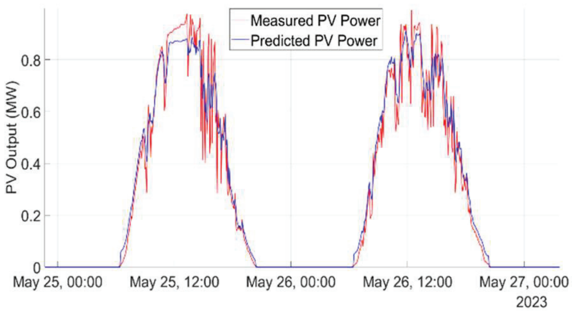

Figure 4. MLPNN 5 Minute DT-Prediction of May 25th-26th, 2023.

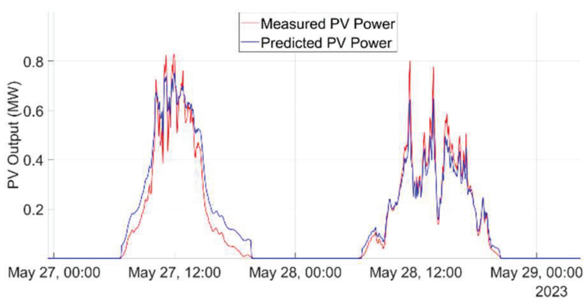

Figure 5. MLPNN 5 Minute DT-Prediction of May 27th-28th, 2023.

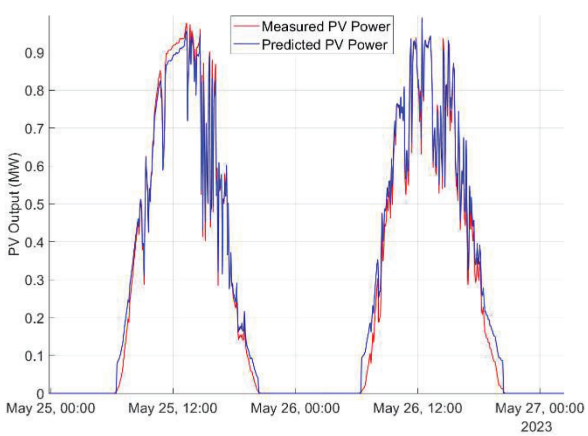

Figure 6. ERNN 5 Minute DT-Prediction of May 25th-26th, 2023.

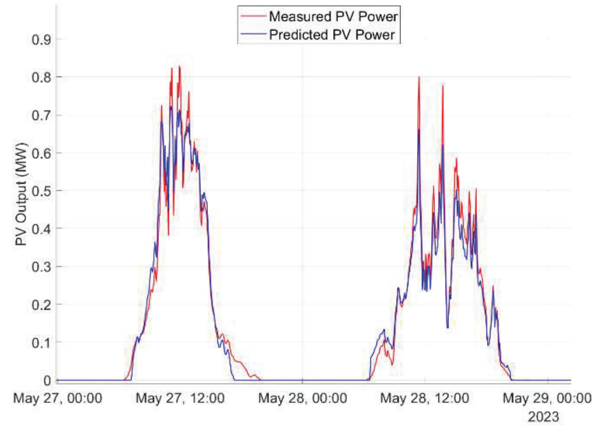

Figure 7. ERNN 5 Minute DT-Prediction of May 27th-28th, 2023.

As can be seen in Figs. 4-7, ERNNs appear to outperform MLPNNs for predictions on a 5-minute time horizon. Table I shows that MLPNNs have lower MAPEs and MSEs than ERNNs for partially cloudy days. However, ERNNs perform better for mostly cloudy days and perform on par with MLPNNs on sunny and moderately cloudy days. This is due to the high volatility of the solar irradiance coupled with the MLPNN's ability to adapt more quickly to change. The different neural networks can be used in an ensemble to generate predictions with higher degrees of accuracy.

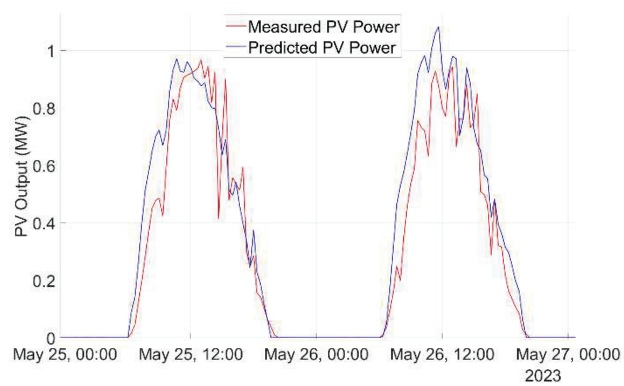

Figure 8. MLPNN 20 Minute DT-Prediction of May 25th-26th, 2023.

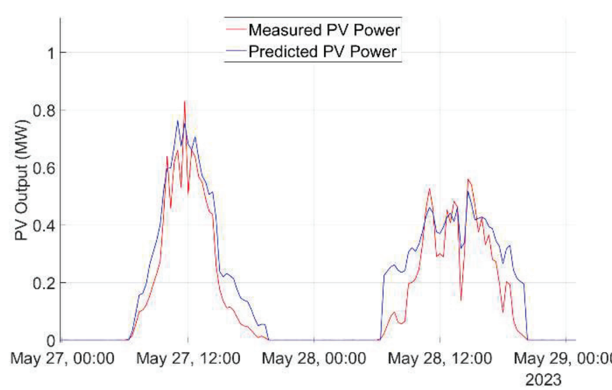

Figure 9. MLPNN 20 Minute DT-Prediction of May 27th-28th, 2023.

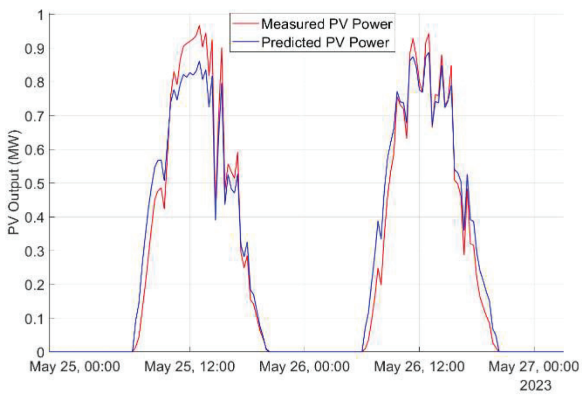

Figure 10. ERNN 20 Minute DT-Prediction of May 25th-26th, 2023.

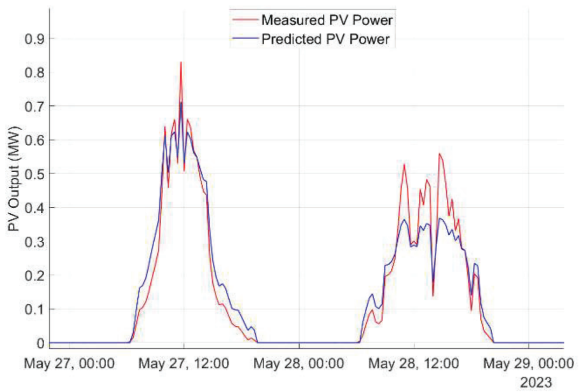

Figure 11. ERNN 20 Minute DT-Prediction of May 27th-28th, 2023.

Performances between MLPNNs and ERNNs are comparable, and they excel in different aspects of the PV profiles. It can be seen from Figs. 8-11 that the ERNNs were able to match the shapes of the curves better than the MLPNNs, but struggled to reach the peaks of the curves. Tables I and II show the daytime mean square error (MSE) and the daytime mean absolute percent error (MAPE) for both 5- and 20-minute predictions.

TABLE I. Neural Network Performance Comparison (5 Minutes)

5-Minute Time Horizon	**MLPNN**		**ERNN**	
Category	***Daytime MSE***	***Daytime MAPE***	***Daytime MSE***	***Daytime MAPE***
Sunny	4.40×10^{-3}	13.90%	7.20×10^{-3}	14.12%
Partly Cloudy	2.29×10^{-2}	25.78%	3.41×10^{-2}	31.62%
Moderately Cloudy	2.53×10^{-2}	30.63%	1.29×10^{-2}	31.22%
Mostly Cloudy	7.60×10^{-3}	23.67%	4.00×10^{-3}	14.05%

TABLE II. Neural Network Performance Comparison (20 Minutes)

20-Minute Time Horizon	**MLPNN**		**ERNN**	
Category	***Daytime MSE***	***Daytime MAPE***	***Daytime MSE***	***Daytime MAPE***
Sunny	9.00×10^{-3}	32.83%	2.95×10^{-2}	60.16%
Partly Cloudy	4.20×10^{-2}	52.51%	4.40×10^{-2}	55.16%
Moderately Cloudy	3.15×10^{-2}	57.16%	2.23×10^{-2}	64.78%
Mostly Cloudy	1.71×10^{-2}	110.06%	1.17×10^{-2}	54.51%

V. Conclusion

In this paper, digital twins of a 1 MW PV plant at Clemson University have been developed for short-term PV power predictions. The DTs are implemented using data from the PV plant using artificial intelligence algorithms. The performances of static and dynamic neural networks based on weather classification and prediction horizon length have been compared. MLPNNs and ERNNs excel in different categories for prediction. Future work for improving the prediction DTs includes investigating distributed neural network architectures such as cellular computational networks. Furthermore, advanced applications in operations and controls of digital twins may be considered as well.

Acknowledgment

The authors acknowledge the facilities provided by the Real-Time Power and Intelligent Systems (RTPIS) Laboratory at Clemson University, SC, USA for the research carried out for this paper.

References

[1] M. AlShafeey, and C. Csáki, "Evaluating Neural Network and Linear Regression Photovoltaic Power Forecasting Models Based on Different Input Methods," Energy Reports, vol. 7, pp. 7601-7614, Nov. 2021

[2] M. Liu, S. Fang, H. Dong, C. Xiu, "Review of Digital Twin about Concepts, Technologies, and Industrial Applications," Journal of Manufacturing Systems, vol. 58 pp. 346-361, Jan. 2021

[3] P. Palensky, M. Cvetkovic, D. Gusain, A. Joseph, (2022, Aug.). Digital twins and their use in future power systems [version 2; peer review: 2 approved]. *Digital Twin 2022.* [Online]. *1*, p. 4. Available: https://digitaltwin1.org/articles/1-4/v2#referee-response-27014

[4] S. Nguyen, M. Abdelhakim and R. Kerestes, "Survey Paper of Digital Twins and their Integration into Electric Power Systems," in Proc. 2021 IEEE Power & Energy Society General Meeting Conf., pp. 01-05.

[5] D. Riley, G. Venayagamoorthy, (2011, July). Characterization and modeling of a grid-connected photovoltaic system using a recurrent neural network. *The 2011 International Joint Conference on Neural Networks.* [Online]. pp. 1761-1766. Available: https://ieeexplore.ieee.org/abstract/document/6033437?casa_token=BI2bOlkBT_kAAAAA:JZ3hLVU2j97bW8XhMVw7fNfHf6r6MCZ6DwIc_iyCVz_T2RPk90YQ01GAtWt4YPRNa5DX9--D

[6] E. VanDerHorn, S. Mahadevan, (2021, June). Digital Twin: Generalization, characterization and implementation. Decision Support Systems. [Online]. 145, Available:

https://www.sciencedirect.com/science/article/pii/S0167923621000348?ref=pdf_download&fr=RR-2&rr=7e4619b569d34581

[7] A. Rasheed, O. San, and T. Kvamsdal. “Digital Twin: Values, Challenges and Enablers from a Modeling Perspective,” IEEE Access, vol. 8, pp. 21980-22012, Jan. 2020.

[8] N. Sehrawat, S. Vashisht, and A Singh, "Solar Irradiance Forecasting Models using Machine Learning Techniques and Digital Twin: A Case Study with Comparison." International Journal of Intelligent Networks, vol. 4, pp. 90-102, May, 2023.

[9] L. Liu, Y. Zhao, D. Chang, J. Xie, Z. Ma, Q. Sun, H. Yin, R. Wennersten. (2018, Oct.). Prediction of short-term PV power output and uncertainty analysis. *Applied Energy.* [Online]. *228*, pp. 700-711. Available: https://www.sciencedirect.com/science/article/pii/S0306261918309826?ref=pdf_download&fr=RR-2&rr=7e46082cd9fa44f3

[10] I. Natgunanathan, V. Mak-Hau, S. Rajasegarar, A. Anwar, (2023, April.). Deakin microgrid digital twin and analysis of AI models for power generation prediction. *Energy Conversion and Management: X.* [Online]. *18*, Available: https://www.sciencedirect.com/science/article/pii/S2590174523000260?ref=pdf_download&fr=RR-2&rr=7e461fe6bb90acdf

[11] R. Kishore, T. Kaur, (2012, June). Backpropagation Algorithm: An Artificial Neural Network Approach for Pattern Recognition, *International Journal of Scientific & Engineering Research*, vol. 3, pp. 1-5, Available: https://www.ijser.org/researchpaper/Backpropagation-Algorithm-An-Artificial-Neural-Network-Approach-for-Pattern-Recognition.pd