Photovoltaic Energy Prediction for New-Generation Cells with Limited Data: A Transfer Learning Approach

Angelo Genovese*, Vera Bernardoni†, Vincenzo Piuri*, Fabio Scotti*, Francesca Tessore‡
* Department of Computer Science, Università degli Studi di Milano, Italy.
{angelo.genovese, vincenzo.piuri, fabio.scotti}@unimi.it
† Department of Physics, Università degli Studi di Milano, Italy. *vera.bernardoni@unimi.it*
‡ Department of Chemistry, Università degli Studi di Milano, Italy. *francesca.tessore@unimi.it*

***Abstract*—Photovoltaic (PV) energy systems are receiving increasing attention, given their relative ease of installation, with 3rd generation technologies promising even simpler fabrication processes and less-intrusive installation possibilities. Therefore, methods for predicting the PV energy output are important to balance the production of other types of renewable sources and avoid wasting energy, with approaches based on machine learning models being especially studied in recent applications. In the case of new-generation cells, limited data is available to train such models, making the use of transfer learning a viable approach to increase prediction accuracy. However, no work in the literature has considered a transfer learning approach studying how much knowledge can be transferred between 2nd and 3rd generation PV technologies. In this paper, we propose the first approach in the literature based on machine learning and transfer learning for the PV energy prediction, in the case of new-generation PV technologies for which limited training data is available. We tested our method on data collected from several locations throughout the world, with results confirming the validity of the approach.**

***Index Terms*—Photovoltaic (PV), Dye-Sensitized Solar Cells (DSSCs), Feedforward Neural Networks (FFNN), Transfer Learning, Renewable Energy**

I. Introduction

Renewable energy sources are gaining increasing attention by the academic and industrial research communities, as well as from the local governments, given the growing concerns towards climate change [1]–[3]. Among renewable energy sources, photovoltaic (PV) energy is one of the most promising and deployed techniques because of its relative ease of deployment and reduced space occupation. In fact, it is possible to install PV systems even in small areas such as personal homes, usually with affordable costs [4], [5]. As an example, Fig. 1 shows the distribution of renewable energies installed in Italy, with PV systems having the greatest share, after hydroelectric systems [6]. However, most PV systems are based on 2nd generation silicon-based PV cells, which have the drawbacks of high production costs and environmental impact [7].

To overcome the disadvantages of 2nd generation PV cells, the research communities are working towards the realization of 3rd generation cells, with dye-sensitized solar cells (DSSCs) emerging as one of the most promising technologies within the new-generation, given their reduced costs and simpler fabrication processes [8], [9]. The widespread use of DSSCs could therefore increase the energy produced using PV systems, expanding their field of application even in areas where traditional silicon-based cells could not be installed (e.g., in absence of ground space or rooftop space), and contributing towards the net zero carbon emission goal [10].

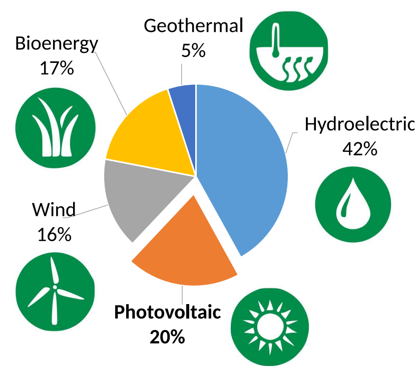

Figure 1. Distribution of renewable energies installed in Italy: photovoltaic (PV) energy has the largest share after hydroelectric energy (source of data: [6]).

In spite of the increasing deployment of PV systems, the energy production may not be enough to account for the whole electricity demand [5]. It is therefore necessary to consider methods for predicting the PV energy produced, in order to plan the production of non-renewable energy accordingly, to avoid wasting energy and storing electrical power, which has a high environmental impact [11]. To this purpose, several techniques have been proposed in the literature, with a particular interest towards the approaches based on machine learning and (recently) deep learning models, which are known to exhibit high accuracies, especially when large quantities of data are available to train them [1], [12].

When new-generation DSSCs cells start to be deployed, there will be the issue of limited available data to train the models, given the shorter amount of operation time with respect to 2nd generation PV cells. To compensate for the problem of limited training data, a typical approach consists in using transfer learning, by pretraining a machine learning model

This work was supported in part by the Università degli Studi di Milano under project 3SUN. We thank the NVIDIA Corporation for the GPU donated.

(e.g., a neural network) on a different but larger database, then tuning the model on a smaller but more specific database [13]. However, transfer learning techniques in the literature usually consider databases belonging to very different application domains (e.g., ImageNet [14]) to pretrain the models. Moreover, no approach in the literature for PV energy prediction has investigated how much knowledge can be transferred between 2nd and 3rd generation cells.

In this paper, we propose an original approach[1] based on machine learning and transfer learning to predict the PV energy production in the case of new-generation cell technologies for which limited training data is available, such as the case for 3rd generation technologies. Our method is based on building a digital twin [15] describing the average PV cell behavior. Then, we calibrate the digital twin by tuning the model on a small target database describing few days of operation of a single kind of PV cell technology, not present in the source database and for which the functioning parameters are not available. To perform the prediction, we consider only meteorological data, rather than historical series of produced power, making our approach location-independent and more generally applicable since it requires only the collection of meteorological measurements, which are publicly available for several locations throughout the world [16]. To the best of our knowledge, this is the first approach in the literature which proposes a transfer learning approach to predict PV energy, based on sharing the knowledge between different PV cell technologies. We tested our approach on data corresponding to different locations in the world, with results confirming the validity of the method.

The paper is structured as follows. Section II reviews the related works. Section III introduces the methodology. Section IV describes the experimental results. Finally, Section V concludes the work.

II. Related Works

It is possible to divide methods for PV energy prediction in two categories, based on the type of data considered: *i)* methods using images (e.g., satellite images, sky images) and *ii)* methods using historical series of numerical data (e.g., meteorological data, numerical weather predictions (NWP), recorded energy output). Another classification divides the methods in three categories based on the type of the approach: *i)* physical methods, *ii)* statistical methods, and *iii)* hybrid methods. In particular, statistical methods include the approaches based on machine learning models (e.g., neural networks) [17], [18]. In this paper, we will focus mainly on the statistical methods based on machine learning and operating with series of numerical data, which have the advantages of not requiring satellite equipment as well as the ability to exploit the large quantities of data that are currently available on the internet [19]. Within these approaches, we will consider the methods that use meteorological data or NWP, rather the ones that need historical series of the produced energy. In fact, methods using meteorological data can be applied to every type of PV system in every location where meteorological data is available. Furthermore, NWP are often the only information available for predicting the energy output, since energy production data may be available only after a delay [1].

Recent methods for PV energy prediction are usually based on artificial neural networks (ANN) or long short-term memory (LSTM) architectures, because of their capability to natively handle tabular data and numerical data in the form of time series. For example, the method described in [20] proposes a methodology based on NWP, ANN, and a statistical postprocessing for the day-ahead forecast of PV energy production. ANNs are also used by the approaches considered in [21], [22], which predict PV energy production 10 minutes in advance, by also merging NWP at a spatial and temporal level. Differently than [20]–[22], which consider ANNs in the form of multilayer perceptrons, the approach proposed in [23] uses ANNs in the form of non-linear auto-regressive neural network with exogenous input (NARX), which are natively designed to process time series.

A method based on LSTM is described in [24], which adds an autoencoder in the architecture, for the purpose of dealing with the uncertainty in NWP. The approach proposed in [25] also uses LSTM to perform the prediction, but rather than using an autoencoder, considers a postprocessing operation based on an integration operator, to increase the reliability of the prediction. While the approaches proposed in [24], [25] aim at increasing the reliability of the prediction using statistical approaches, the method described in [17] integrates the LSTM with a physical model, with the purpose of incorporating domain knowledge into the architecture and therefore reducing the need for large quantities of training data.

Lastly, there are some approaches based on machine learning that explicitly consider transfer learning [4], [13], however to the best of our knowledge no method in the literature has yet considered a transfer learning between different PV technologies.

III. Methodology

We propose an innovative approach based on machine learning and transfer learning for predicting the energy output of PV cells when only limited training data is available for that specific PV technology and the functioning parameters are not available, for example in the case of 3rd generation cells. In our approach, we consider only meteorological data as input for the prediction.

The methodology is composed of the following steps: *i)* database creation; *ii)* feature extraction; *iii)* creation of the digital twin; *iv)* transfer learning: calibration of the digital twin. Fig. 2 outlines the proposed methodology.

A. Database Creation

We define a set S of locations for which to extract meteorological data, $S = \{l_1, l_2, \ldots, l_M\}$, where l_i is the i-th location and M is the number of locations considered. Each location

[1]The source code will be available upon acceptance of the paper at: `http://iebil.di.unimi.it/PVNet/index.htm`

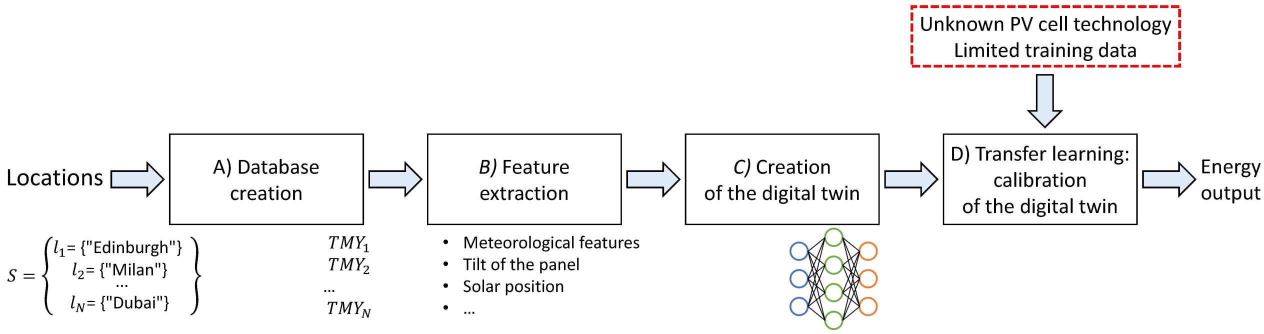

Figure 2. Outline of the proposed methodology for predicting the energy output of PV cells when only limited training data is available for that specific PV technology and the functioning parameters are not available, for example in the case of 3rd generation cells.

is defined by its GPS coordinates, its name, and the elevation. For example, the location l_1:

$$l_1 = \{55.94, -3.14, \text{“Edinburgh”}, 70\} \tag{1}$$

is the city of Edinburgh, located at latitude $55.94°N$, longitude $3.14°W$, and elevation 70 m.

For each location l_i, we use the PVLib software [26] to extract the typical meteorological year (TMY) from the photovoltaic geographical information system (PVGIS) [16]. The TMY consists of meteorological measurements for a particular location, with a temporal resolution of 1 h. In particular, at each instant of time $\mathbf{t}$, defined as $\mathbf{t} = [\text{year}, \text{month}, \text{day}, \text{hour}]$, the TMY includes the following set $M(\mathbf{t})$ of meteorological features:

$$M(\mathbf{t}) = \begin{bmatrix} \text{temp\_air } [°C], \\ \text{relative\_humidity } [\%], \\ \text{ghi } [W/m^2], \\ \text{dni } [W/m^2], \\ \text{dhi } [W/m^2], \\ \text{IR(h) } [unitless], \\ \text{wind\_speed } [m/s], \\ \text{wind\_direction } [°], \\ \text{pressure } [Pa] \end{bmatrix}, \tag{2}$$

where “ghi” is the global horizontal irradiance, “dni” is the direct normal irradiance, and “dhi” is the diffuse horizontal irradiance.

The proposed method considers data with 1 h time granularity. It is possible to perform predictions with higher time granularity, especially in geographical areas where weather conditions may change rapidly, however accumulation systems present in modern PV systems can usually compensate for rapid fluctuations in photovoltaic energy output [27].

The TMY_i for each location i is then composed by the stacked vectors of meteorological features $M(\mathbf{t})$, extracted for each instant of time $\mathbf{t}$, with 1 h interval. For example, the TMY_i for the i-th location and for the year 2005 is composed as follows:

$$\text{TMY}_i = \begin{bmatrix} M(\mathbf{t}_1 = [2005, 1, 1, 00{:}00]), \\ M(\mathbf{t}_2 = [2005, 1, 1, 01{:}00]), \\ \dots \\ M(\mathbf{t}_T = [2005, 12, 31, 23{:}00]), \end{bmatrix}, \tag{3}$$

where $\mathbf{t}_1 = [2005, 1, 1, 00{:}00]$ represents the first hour of the year 2005, at 00:00 on January 1st, 2005, and $\mathbf{t}_T = [2005, 12, 31, 23{:}00]$ represents the last hour of 2005, at 23:00 on December 31st, 2005.

B. Feature Extraction

We extract the features by using the TMY for each location and compute the information related to the tilt of the panel, solar position, extraterrestrial radiation, pressure, incidence, irradiance, and cell temperature. We consider such features since they can be computed using meteorological information and are directly related to the energy output of the PV cell [28]. Table I lists the complete feature set. Moreover, the PVLib [26] functions used to compute the features are listed in the website of the project[1].

Meteorological features (M) *(1–9)*. The set of meteorological features M, computed as described in Section III-A.

Tilt of the Panel (10). We compute it as the latitude of the location. For example, at a latitude of $45°N$ the tilt is 45°.

Solar Position (11–16). Using the GPS coordinates, the altitude, the temperature, and the air pressure, we compute a set of features describing the actual and apparent position of the sun.

Extraterrestrial radiation (17). We compute it based on the day of the year.

Pressure features (18–20). Using the altitude, solar position, and pressure, we compute the relative and absolute airmass.

Incidence (21). We compute the angle of incidence on the PV cell by using the information of the solar position.

Irradiance (22–26). We use the solar position, the global horizontal irradiance, the direct normal irradiance, the diffuse horizontal irradiance, and the extraterrestrial radiation to compute a set of features describing the irradiance.

Cell temperature (27). It is computed by using the irradiance, the temperature of the air, and the wind speed.

We compute the described features for each instant of time $\mathbf{t}$ present in the TMY. As a result, we obtain the feature set F with size $T \times 27$, where T is the number of hourly samples in the TMY.

C. Creation of the Digital Twin

We create a digital twin that describes the average PV cell behavior, by training a neural model using the features extracted as described in Section III-B and the ground truth computed considering several kinds of PV cell technologies.

Table I
FEATURE SET USED IN THE PROPOSED METHODOLOGY FOR PREDICTING THE PV ENERGY.

Type	**N.**	**Name**
Meteorological features (M)	1	temp_air $[°C]$
	2	relative_humidity $[\%]$
	3	ghi $[W/m^2]$
	4	dni $[W/m^2]$
	5	dhi $[W/m^2]$
	6	IR(h) $[unitless]$
	7	wind_speed $[m/s]$
	8	wind_direction $[°]$
	9	pressure $[Pa]$
Tilt of the panel	10	surface_tilt $[°]$
Solar position	11	apparent_zenith $[°]$
	12	zenith $[°]$
	13	apparent_elevation $[m]$
	14	elevation $[m]$
	15	azimuth $[°]$
	16	equation_of_time $[°]$
Extraterrestrial radiation	17	dni_extra $[W/m^2]$
Pressure features	18	airmass $[unitless]$
	19	pressure $[Pa]$
	20	am_abs $[unitless]$
Incidence	21	aoi $[°]$
Irradiance	22	poa_global $[W/m^2]$
	23	poa_direct $[W/m^2]$
	24	poa_diffuse $[W/m^2]$
	25	poa_sky_diffuse $[W/m^2]$
	26	poa_ground_diffuse $[W/m^2]$
Cell temperature	27	cell_temperature $[°C]$

First, we extract the ground truth information for each instant of time $\mathbf{t}$ by considering the performance model described in [28]. Such model associates the feature set at each instant of time $F(\mathbf{t})$ to the corresponding energy output of the PV cell $G(\mathbf{t})$, considering the parameters of each PV cell technology. The PVLib [26] functions used to extract the ground truth are listed in the website of the project[1].

We extract the ground truth information considering a set of 10 PV cell technologies, with each j-th PV cell technology defined by its set of parameters C_j. We chose the PV technologies and the inverters by randomly sampling the ones available in PVLib. As a result, we obtain the ground truth information $G(\mathbf{t}, C_j)$ for each time $\mathbf{t}$ and for each PV cell technology C_j.

Second, to create the digital twin, we consider a machine learning model based on feedforward neural networks (FFNN) and trained using F as input and G as the target. The purpose is to build a neural model that learns the association between the considered features, which only depend from meteorological information, and the energy output of the PV system. In particular, we consider the energy output produced using multiple PV cell technologies to obtain a digital twin of the PV cell in the form of a generalized predictor, which has learned the average PV cell behavior (Fig. 3).

D. Transfer Learning: Calibration of the Digital Twin

We propose a procedure based on transfer learning to calibrate the digital twin and then use it to perform the prediction of PV energy in the case of a PV cell technology

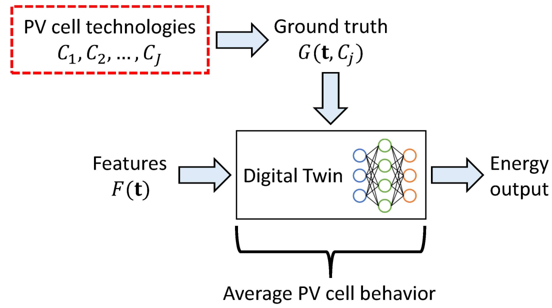

Figure 3. Digital twin of the PV cell, which has learned the average PV cell behavior in associating the considered features F, which only depend from meteorological information, to the ground truth of energy output G.

for which limited training data is available and the functioning parameters C are not available, such as the case of 3rd generation cells.

First, to simulate a situation in which it is necessary to predict the energy output of a PV cell for which limited training data is available and for which the parameters C are not available, we select a portion of the data (F, G) that respects both the following conditions: *i)* corresponds to a period of time not present in the data used to create the digital twin; *ii)* corresponds to the energy output of a PV cell technology that was not present in the data used to create the digital twin.

Second, we perform a calibration of the digital twin by tuning the neural model on the selected portion of the data. In particular, we consider a deep tuning approach, a transfer learning procedure in which all the weights of the neural model are updated during the learning process.

IV. EXPERIMENTAL RESULTS

A. Database

We consider the following set S of locations to create the database and extract the TMY, as described in Section III-A:

$$S = \left\{ \begin{array}{l} l_1 = \{55.94, -3.14, \text{“Edinburgh”}, 70\} \\ l_2 = \{45.46, 9.19, \text{“Milan”}, 134\} \\ l_3 = \{45.05, 9.68, \text{“Piacenza”}, 63\} \\ l_4 = \{38.11, 13.35, \text{“Palermo”}, 30\} \\ l_5 = \{25.26, 55.31, \text{“Dubai”}, 2\} \end{array} \right\} \quad (4)$$

For each location $l_i \in S$, we extracted the features for all the years in the range $(2000, 2020)$, as described in Section III-B.

B. Validation Procedure and Error Measures

To evaluate the capability of the proposed approach to predict the energy output for PV cell technologies for which limited data is available and the functioning parameters are not available, we proposed a specific year-fold/PV-fold cross-validation procedure, shown in Fig. 4 and composed of four steps:

1) We extract the testing data F_{test}, G_{test} by considering a single year and a single PV cell technology, while we extract the training data F_{train}, G_{train} by considering the remaining data.
2) We create the digital twin as described in Section III-C, by training the neural model using F_{train}, G_{train}.

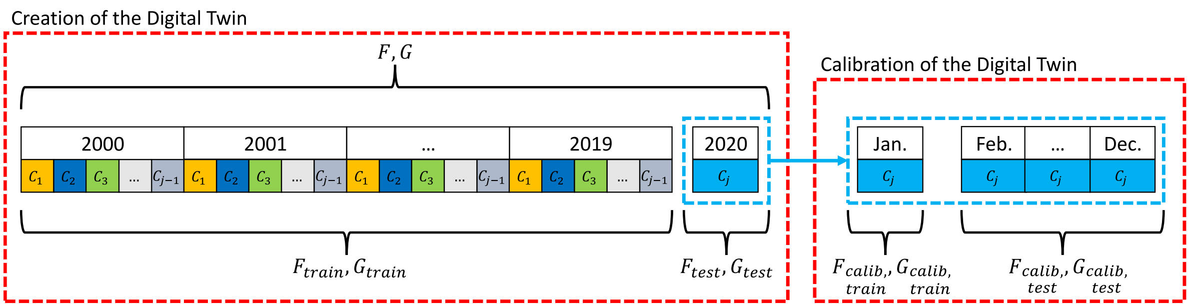

Figure 4. Outline of the proposed year-fold/PV-fold cross-validation procedure. For creating the digital twin, we divide the data by excluding a single year (e.g., 2020) and a single PV cell technology (e.g., C_j), to obtain the training data F_{train}, G_{train}. Then, for calibrating the digital twin, we further divide the testing data by selecting the data corresponding to a single month (e.g., January), to obtain the training data for the calibration step $F_{calib,train}, G_{calib,train}$. In this way, we aim to simulate the prediction of the energy output for PV cell technologies for which limited data is available and the functioning parameters are not available.

3) We further divide the testing data F_{test}, G_{test} to extract the training data for the calibration step $F_{calib,train}, G_{calib,train}$, by selecting the data corresponding to a single month, and the testing data $F_{calib,test}, G_{calib,test}$ for the calibration step as the remaining data.
4) We calibrate the digital twin as described in Section III-D, using $F_{calib,train}, G_{calib,train}$ and we test it using $F_{calib,test}, G_{calib,test}$.

We repeat the procedure by looping on all the years in the range $(2000, 2020)$ and on all the 10 different PV cell technologies considered, then we average the results.

As error measure, we consider the mean absolute error (MAE), defined as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |p_i - \hat{p}_i| \quad , \tag{5}$$

where p_i is the real energy output and $\hat{p}_i$ is the energy output predicted by the calibrated digital twin.

C. Neural Architecture and Learning

We consider a digital twin modeled using a fully-connected FFNN, with an input layer composed by 27 nodes, a hidden layer with 50 nodes, and an output layer with 1 node. The hidden layer has a sigmoidal transfer function, while the input and output layers have a linear transfer function. The choice of the neural architecture is made purely as an application example and different machine learning approaches can be considered.

During both the creation and calibration of the digital twin we train the FFNN using the SGD algorithm for 10 epochs, with a batch size of 50 samples and a learning rate $lr = 2e^{-4}$. Before training, we perform a min-max normalization of the data, with each feature normalized separately.

D. Accuracy

We evaluate the accuracy of the proposed method using the validation procedure described in Section IV-B, using $F_{calib,test}, G_{calib,test}$ as test data. We compare the results obtained both without performing the calibration and with performing the calibration of the digital twin (*proposed approach*).

Table II
Average accuracy of the proposed methodology in terms of mean absolute error (MAE).

Location	MAE	
	Without calibration	With calibration (*proposed approach*)
Edinburgh	22.74	**14.56**
Milan	33.03	**13.94**
Piacenza	28.93	**13.57**
Palermo	30.92	**13.48**
Dubai	37.30	**13.98**

Table II reports the average accuracies obtained for the different locations, while Fig. 5 shows examples of predictions for different locations. From the Table and the Figure, it is possible to observe that the proposed approach, with the calibration of the digital twin, obtains the best results in terms of MAE.

The results confirm the validity of the methodology, indicating that the proposed digital twin can take advantage of the data generated using several configurations of PV cell technologies to learn an average behavior of the PV cells, which can then be calibrated with limited effort using training data captured within a short amount of time with the target PV technology.

To evaluate the robustness of our approach to varying levels of data quality, we perform a sensitivity analysis by adding noise to $F(\mathbf{t})$ and to $G(\mathbf{t})$. We add the noise separately for each meteorological feature, by considering additive Gaussian noise with mean $m = 0$ and standard deviation σ equal to a percentage $p < 1$ of the average feature value $\sigma = p \cdot \overline{f}$, with $p \in \{0.05, 0.1, 0.2\}$. Table III reports the corresponding accuracies. From the Table, it is possible to observe that the proposed approach enables to decrease the MAE even in the presence of noisy data.

V. Conclusions

This paper presented the first approach in the literature for the prediction of the energy output of new-generation PV cells in the case of limited training data. The approach is based on creating a digital twin, which consists in a FFNN predictor trained using the parameters of several different PV cell technologies, then calibrating the digital twin with an approach based on transfer learning, by tuning the FFNN predictor using data corresponding only to the new-generation PV cell and describing a limited time period of a single month.

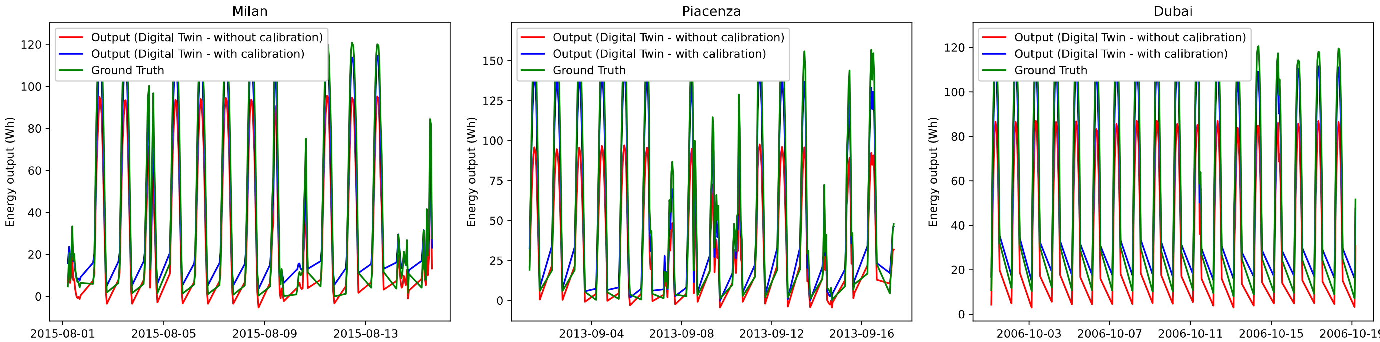

Figure 5. Examples of PV energy output predictions for different locations, obtained both without performing the calibration and with performing the calibration of the digital twin (*proposed approach*). It is possible to observe that the proposed approach with the calibration of the digital twin (*blue*) obtains the prediction closest to the ground truth (*green*).

Table III
AVERAGE ACCURACY OF THE PROPOSED METHODOLOGY IN THE PRESENCE OF ADDITIVE GAUSSIAN NOISE.

Location	MAE $\mathbf{p = 0.05}$ W/out calib.	With calib.	$\mathbf{p = 0.1}$ W/out calib.	With calib.	$\mathbf{p = 0.2}$ W/out calib.	With calib.
Edinburgh	22.42	16.01	22.16	15.83	22.89	15.65
Milan	33.58	16.69	33.48	17.92	34.58	19.65
Piacenza	27.71	16.70	27.19	17.90	28.39	18.17
Palermo	31.25	16.07	31.16	17.22	31.76	18.15
Dubai	37.60	15.81	37.63	16.84	37.29	17.92

We evaluated the approach using a specific year-fold/PV-fold cross-validation procedure, with results confirming the validity of our methodology, by showing an increased prediction accuracy with respect to a non-calibrated digital twin, even when only a very limited data is used during the calibration step. To further increase the accuracy, future works will consider different neural architectures and databases, as well as data with higher time granularity.

REFERENCES

[1] R. Donida Labati, A. Genovese, V. Piuri, F. Scotti, and G. Sforza, "A decision support system for wind power production," *IEEE Trans. Syst., Man, Cybern. Syst*, vol. 50, no. 1, 2020.

[2] "UN climate change conf. of the parties (COP26)," 2021. [Online]. Available: https://ukcop26.org/

[3] "The Paris Agreement," 2015. [Online]. Available: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement

[4] H. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Wei, and G. Sun, "Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning," *Int. J. Electr. Power Energy Syst.*, vol. 118, 2020.

[5] M. Lazzaroni, S. Ferrari, V. Piuri, A. Salman, L. Cristaldi, and M. Faifer, "Models for solar radiation prediction based on different measurement sites," *Measurement*, vol. 63, 2015.

[6] "Renewable energy in Italy," 2021. [Online]. Available: https://www.enelgreenpower.com/learning-hub/renewable-energies/italy/

[7] G. Di Carlo, A. O. Biroli, F. Tessore, S. Caramori, and M. Pizzotti, "β-substituted Zn^{II} porphyrins as dyes for DSSC: A possible approach to photovoltaic windows," *Coord. Chem. Rev.*, vol. 358, 2018.

[8] B. O'Regan and M. Grätzel, "A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO_2 films," *Nature*, vol. 353, 1991.

[9] M. Freitag, J. Teuscher, Y. Saygili, X. Zhang, F. Giordano, P. Liska, J. Hua, S. M. Zakeeruddin, J.-E. Moser, M. Grätzel, and A. Hagfeldt, "Dye-sensitized solar cells for efficient power generation under ambient lighting," *Nature Photon.*, vol. 11, 2017.

[10] IEA - International Energy Agency, "Net Zero by 2050 - A Roadmap for the Global Energy Sector," Tech. Rep., 2021. [Online]. Available: https://www.iea.org/reports/net-zero-by-2050

[11] R. Donida Labati, A. Genovese, V. Piuri, and F. Scotti, "Towards the prediction of renewable energy unbalance in smart grids," in *Proc. of RTSI*, 2018.

[12] Y. Cai, A. Genovese, V. Piuri, F. Scotti, and M. Siegel, "IoT-based architectures for sensing and local data processing in ambient intelligence: Research and industrial trends," in *Proc. of I2MTC*, 2019.

[13] S. Zhou, L. Zhou, M. Mao, and X. Xi, "Transfer learning for photovoltaic power forecasting with Long Short-Term Memory neural network," in *Proc. of BigComp*, 2020.

[14] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. of CVPR*, 2009.

[15] E. Negri, L. Fumagalli, and M. Macchi, "A review of the roles of digital twin in CPS-based production systems," *Procedia Manuf.*, vol. 11, 2017.

[16] EU Science Hub, "Photovoltaic Geographical Information System (PVGIS)," 2021. [Online]. Available: https://ec.europa.eu/jrc/en/pvgis

[17] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, 2021.

[18] L. Stoyanov, Z. Zarkov, I. Draganovska, and V. Lazarov, "Methods for energy production estimation from photovoltaic plants: Review and application," in *Proc. of HiTech*, 2018.

[19] J. T. Dellosa and E. C. Palconit, "Artificial Intelligence (AI) in renewable energy systems: A condensed review of its applications and techniques," in *Proc. of EEEIC / I CPS Europe*, 2021.

[20] S. Theocharides, G. Makrides, A. Livera, M. Theristis, P. Kaimakis, and G. E. Georghiou, "Day-ahead photovoltaic power production forecasting methodology based on machine learning and statistical post-processing," *Appl. Energy*, vol. 268, 2020.

[21] F. Rodríguez, F. Martín, L. Fontán, and A. Galarza, "Ensemble of machine learning and spatiotemporal parameters to forecast very short-term solar irradiation to compute photovoltaic generators' output power," *Energy*, vol. 229, 2021.

[22] F. Rodríguez, A. Galarza, J. C. Vasquez, and J. M. Guerrero, "Using deep learning and meteorological parameters to forecast the photovoltaic generators intra-hour output power interval for smart grid control," *Energy*, vol. 239, 2022.

[23] M. A. Hassan, N. Bailek, K. Bouchouicha, and S. C. Nwokolo, "Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks," *Renew. Energy*, vol. 171, 2021.

[24] Y. Zhang, C. Qin, A. K. Srivastava, C. Jin, and R. K. Sharma, "Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model," *IEEE Trans. Ind Appl.*, vol. 56, no. 6, 2020.

[25] M. Abdel-Nasser, K. Mahmoud, and M. Lehtonen, "Reliable solar irradiance forecasting approach based on choquet integral and deep LSTMs," *IEEE Trans. Ind. Informat.*, vol. 17, no. 3, 2021.

[26] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *J. Open Source Softw.*, vol. 3, no. 29, 2018.

[27] V. Benda, "Photovoltaics, including new technologies (thin film) and a discussion on module efficiency," in *Future Energy (Third Edition)*, T. M. Letcher, Ed. Elsevier, 2020.

[28] D. L. King, W. E. Boyson, and J. A. Kratochvil, "Sandia photovoltaic array performance model," Sandia National Laboratories, Albuquerque, NM, Tech. Rep. 3535, 2004.