Effectiveness of neural networks and transfer learning to forecast photovoltaic power production

Applied Soft Computing Journal 149 (2023) 110988

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Effectiveness of neural networks and transfer learning to forecast photovoltaic power production

Andrea Bellagarda [a], Donato Grassi [a], Alessandro Aliberti [c,*], Lorenzo Bottaccioli [b,c], Alberto Macii [a], Edoardo Patti [a,b]

[a] Department of Control and Computer Engineering, Politecnico di Torino, 10129 Torino, Italy
[b] Energy Center Lab, Politecnico di Torino, 10129 Torino, Italy
[c] Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, 10129 Torino, Italy

ARTICLE INFO

MSC:
00-01
99-00

Keywords:
Transfer learning
Artificial Neural Networks
Photovoltaic energy production forecasting
Smart building
Energy efficiency
Demand side management

ABSTRACT

Artificial Neural Networks (ANNs) can successfully be integrated into smart models for energy prediction, but require large datasets for training. This investigation presents an innovative methodology for photovoltaic power generation forecasting with ANNs, when only a limited amount of real data is available, and has been tested and validated on a real-life photovoltaic installation. Feature selection identifies which meteorological features most impact photovoltaic power generation. A simulator, which accurately models a real photovoltaic installation, is used to create an artificial, but accurate and realistic, dataset of power generation large enough to effectively train and test different ANNs. These are then exploited on a portion of real, but limited, dataset of power generated by the real photovoltaic installation on which the simulator is modeled. Finally, different transfer learning techniques are used to tune the ANN models with the remaining portion of the real, but limited, dataset of photovoltaic power generation.

1. Introduction

As highlighted by the International Energy Agency (IEA), in its 2022 World Energy Outlook Report [1], with the world in the midst of a global energy crisis, and faced with energy shortfalls and high prices, governments have rushed to try and secure alternative energy sources and supplies, and accelerating the flow of new renewables projects. Within this context, electricity accounts for about 20% of the world's total final consumption of energy, but its share of energy services is higher. Investments in clean electricity and electrification, along with expanded and modernized grids, offer clear and cost-effective opportunities to cut emissions more rapidly while bringing electricity costs down from their current highs. In the most affected regions of this energy crisis, for example, higher shares of renewables were correlated with lower electricity prices, and more efficient homes and electrified heat have provided an important buffer for some – but far from enough – consumers. Today's growth rates for deployment of solar photovoltaic (PV) and wind, if maintained, would lead to a much faster transformation than projected in the Stated Policies Scenarios (STEPS), although this would require supportive policies not just in the leading markets for these technologies but across the world. Supply chains for some key technologies – including solar PV – are expanding at rates that support higher global ambition. If all announced manufacturing expansion plans for solar PV see the light of day, manufacturing capacity would exceed the deployment levels in the Announced Pledges Scenario (APS) in 2030 by around 75% and approach the levels required in the Net Zero Emissions (NZE) Scenario. These clean energy supply chains are a huge source of employment growth, with clean energy jobs already exceeding those in fossil fuels worldwide and projected to grow from around 33 million today to almost 55 million in 2030 in the APS. Finally, it is interesting to underline how the World Energy Outlook Report also highlighted that demand-side measures have generally received less attention, but greater efficiency is an essential part of the short- and longer-term response.

Within this framework, it is clear how the role of renewable energy sources such as PV energy is becoming increasingly important. PV energy can be classified in the variable renewable energy (VRE) sources, due to its fluctuating power output based on solar power. The variable nature of PV power poses a challenge towards its use as a reliable energy source in power grids, whose stability strongly depends on the balance between generation and consumption of energy. According to the IEA, a power system is flexible, if it can, within economic boundaries, respond quickly to high fluctuations in supply and demand,

* Corresponding author.
E-mail address: alessandro.aliberti@polito.it (A. Aliberti).

https://doi.org/10.1016/j.asoc.2023.110988
Received 17 April 2023; Received in revised form 18 October 2023; Accepted 26 October 2023
Available online 30 October 2023

Acronyms

1D-CNN	1 Dimensional Convolutional Neural Network
ADAM	Adaptive Moment Estimation
ANN	Artificial Neural Network
ANNs	Artificial Neural Networks
APS	Announced Pledges Scenario
BiLSTM	Bidirectional LSTM
C-LSTM	Constrained LSTM
CLSTM	Contextual LSTM
CNN	Convolution Neural Network
GHI	global horizontal irradiance
GRU	Gated Recurrent Unit
IEA	International Energy Agency
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-Term Memory
MAD	Mean Absolute Difference
MAPE	Mean Absolute Percentage Error
MI	mutual information
MSE	Mean Squared Error
NAR	Non-linear AutoRegressive
NZE	Net Zero Emissions
PV	photovoltaic
R^2	Coefficient of Determination
RMSD	Root Mean Square Difference
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SBS	Sequential Backward Selection
SFS	Sequential Forward Selection
STEPS	Stated Policies Scenarios
TL	Transfer Learning
TW	Time Window
UV	ultraviolet
VRE	variable renewable energy

ramping down a generation when demand decreases, and upwards when it increases for scheduled and unpredictable events [2]. However, the flexibility of power systems has become a concept that needs to be redefined, due to the increasing penetration levels of power generation from variable and hardly predictable sources such as wind and solar energy, which generate uncertainty on the supply side [3].

These difficulties however can be overcome with the support of technological innovations such as smart grids, which allow to improve management and stability of existing power grids by integrating them with modern distributed computing facilities and communication networks [4]. Within the smart grid framework, innovative applications can be implemented in order to better coordinate power demand and supply, for example real time forecasting or demand response [5], which can adapt the power consumption in order to align power demand with supply. For such applications to work effectively though, the ability to effectively forecast both power demand and supply with different forecast horizon has become increasingly important. Within this framework, specific attention can be given to the effective forecasting of PV power generation [6].

According to [7,8] and [9], artificial intelligence techniques have become an excellent tool for wind generation and PV generation forecasting. [10] discuss how forecasting based on Artificial Neural Network (ANN), which have been used in a variety of fields since [11] first published his comprehensive book of many neural network techniques from an engineering perspective, is one of the most effective methods for PV generation forecasting. However, [10] also highlight some main drawbacks of ANNs such as the fact that they require a large amount of data for the training process, a random initial dataset that may reduce the reliability of the forecasted results, and the challenge and time required for accurate development of the model architecture. Nonetheless, different ANN solutions have been investigated for long-, medium- and short-term PV generation forecasting.

For long-term prediction horizons, [12] used different ways to forecast PV power from 0 h to 48 h ahead, based on spatial clustering of the PV fleet and an ensemble of multilayer perceptron using satellite and numerical weather prediction data. For mid-term prediction horizons, [13] and [14] investigate different approaches for forecasting of PV power generation up to 24 h ahead, each using different ANNs and meteorological inputs, with results showing how the latter increase the prediction performance. [15] also investigate forecasting PV power generation up to 24 h ahead, using PV simulation software to generate the data inputs to their model, with results showing how their generalized model could effectively forecast PV power generation in both normal (clear-sky) and abnormal days (cloudy or rainy days), and also in various seasons and different abnormal weather conditions. For shorter prediction horizons, [16] investigate one step and multi step ahead PV power generation forecasting for 1 min, 5 min, 30 min and 60 min, by using Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), Bidirectional GRU, Convolution Neural Network (CNN), and hybrid architecture. Different model architectures are tested and their performance compared to that of a Non-linear AutoRegressive (NAR) neural network and a Elman recurrent neural network. Results show that the best performance is obtained with LSTM and GRU based models, with LSTM presenting the best accuracy and simplicity of implementation.

After the main LSTM paper was published in the journal Neural Computation in 1997 [17], this ANN has been used successfully in a variety of fields, and after it started showing interesting performance when applied to PV power generation forecasting, different authors investigated its performance in this field. [18] experiment with different LSTM model parameters (number of hidden nodes, activation function, number of input variables), and also by changing the division of the dataset, with their results showing good performance in forecasting the daily PV power generation, while again also highlighting how increasing the number of input variables does not produce better results. [19], propose three different models for PV power forecast: CNN, LSTM and Contextual LSTM (CLSTM). Their results shows that all of the models have a positive performance, with CLSTM outperforming the others in terms of accuracy while GRU presents the shortest training time. Finally, [20] also present an investigation leveraging neural networks for PV power prediction, including GRU, which provides the best prediction accuracy.

[21] instead investigate PV power forecasting using the global horizontal irradiance (GHI) measured by sensors and GHI in clear sky condition. The authors evaluate the effectiveness of exogeneous inputs together with different machine learning models (Feedforward, Echo State, 1D-Convolutional, Long Short-Term Memory neural networks and Random Forest) for short term solar radiation forecasting. The investigation's results show that best model is GRU, and that exogenous inputs significantly improve the forecasting performance for prediction horizons greater than 15 min, while for very short prediction horizons (i.e. 15 min) the improvements are negligible.

Regardless the range of prediction, one of the main challenges posed techniques belonging to the field of machine learning and artificial intelligence, particularly those related to ANNs, is that they require significant amounts of data in order to be trained effectively and produce acceptably accurate results. [19], for example, in their work with CNN, GRU and CLSTM, recommend to select a data length of at least 3 years.

In recent years, Transfer Learning (TL) has been increasingly investigated as a possible solution to overcome the challenge posed by lack of large and reliable enough data in a variety of fields. After [22] published the first paper addressing Transfer Learning (TL) in neural network training, this methodology has been investigated and applied in a variety of fields. Its application has also been investigated in PV power forecasting but, to the best of our knowledge, research availability on such topic is still scarce.

[23], for example, propose a method to transfer the knowledge obtained with historical solar irradiance data to the PV power output prediction, using an GRU model which is first trained using historical solar irradiance data, and then fine-tuned with the PV output data. The author use 6 months of historical solar irradiance data and 45 days of historical output data, with a sampling interval of 10 min. The results show that although the amount of training data is insufficient to generate accurate results, TL does improve the performance of the prediction, with Mean Absolute Percentage Error (MAPE) improving an average of 23% and Root Mean Square Error (RMSE) improving an average of 10%, for predictions up to 40 min ahead.

In another investigation, [24] focus on newly-constructed PV plants and on their challenge to effectively execute hourly day-ahead PV power generation, by applying a Constrained LSTM (C-LSTM) model together with two parameter-transferring strategies, thus combining TL and deep learning models. The results show how standard GRU models are outperformed by C-LSTM models in terms of forecasting accuracy, and also how variability and accuracy issues due to different sky conditions can be improved by the application of the proposed combination of C-LSTM and TL strategies.

[25] also developed a model based on TL to predict PV power generation, with their experimental results showing how traditional learning methods are outperformed by the proposed TL model. The model they developed leverages the variation of solar altitude angles in a year to identify the season, combine it with the meteorological factors hidden in the data collected from a PV system, and use it as input for online learning models based on both traditional learning and TL approaches to predict power generation.

Within this framework, the purpose of this investigation is therefore to present an innovative methodology for PV power generation forecasting with ANNs, when only a limited amount of real data is available. Feature selection is first used to investigate different meteorological features, such as GHI, humidity, air temperature, and so on, in order to identify those which most impact the accuracy of data prediction forecast. The PV power generation simulator presented by [26], which is accurately modeled to replicate real PV installations, is then used. As a case study, we selected a PV system installed in the rooftops of some buildings in our University campus, located in Turin, north west of Italy. Thus, the PV power generation simulator has been applied to these rooftops to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. Moreover, for these real world PV system, we also collected measurements of real power generated, as a second dataset. It is worth noting that real and simulated PV installations are coincident. In the proposed methodology, the simulated dataset, together with the meteorological features previously selected, is used for the initial training and testing of the ANNs. The resulting ANN models, trained and tested on the simulated dataset, are then exploited on a portion of the real dataset to evaluate their prediction performance against real data. Finally, different TL techniques are used to tune the ANN models with the remaining portion of the real dataset investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. As already mentioned, the whole methodology has been tested and validated on a real-life PV installation in our university campus.

The novelty of this investigation lies in the exploitation of a PV power generation simulator, which accurately models a real PV installation, to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANN models, which are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data. Further novelty is brought by the application of different TL techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. The ANNs used in this investigation are: GRU, which is a well-established ANNs in PV power generation forecasting; and 1 Dimensional Convolutional Neural Network (1D-CNN), which is a variation of CNN, which have occasionally been used in PV power generation forecasting.

The results of the investigation conclude that after the feature selection process, the variables which most impact power prediction are global horizontal irradiance, humidity, temperature, dew point, ultraviolet index, sunshine duration and time of the day; the best prediction performance, up to 4 h, is given by GRU neural networks; the TL techniques can successfully improve short-term forecasting performance up to 2 h.

The rest of this manuscript is organized as follows. Section 2 introduces the case study. Section 3 presents the proposed methodology. Section 4 discusses our experimental results. Finally, Section 5 provides our concluding remarks.

Table 1
Technical specifications of the PV system.

Installation year	2016	
Size	631 kWp	
# of PV cells	96	
Power PV cells	3406 Wp	
Brand PV cells	BEN-Q	
# of PV modules	1836	
BRAND PV modules	BENQ SOLAR	
MODEL PV modules	SunForte PM096B00	
Modules' technology	MONO-CRISTALLINO	
Modules' nominal power	327 W	
Modules' surface area	1.63 m^2	
# of inverters	8	19
Max power inverters	20 kW	25 kW
Brand inverters	SMA	
Modelinverters	Sunny tripower	

2. Case study

The methodology presented in this work, aiming to forecast PV power generation with ANNs, when only a limited amount of real data is available, tested and validated on a real-life PV installation located on the rooftop of a building of our university campus in Turin, Italy, as shown in Fig. 1.

The PV installation has a total surface area of almost 3000 m^2, and the characteristics described in Table 1.

The PV power production of this installation is measured through specific sensors, with data collection occurring every 15 min. Data collected through these sensors covers years 2018 to 2020, for a total of 105,216 data points.

The PV power generation simulator presented by [26], which accurately replicates the real PV installation located on the roof of our university campus in Turin, Italy, is then used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. This artificial, but accurate and realistic dataset, together with the meteorological features previously selected, is used for the initial training and testing of the ANNs. The simulated dataset also presents PV power production every 15 min. The simulator was used to generate an artificial dataset covering years 2010 to 2015, for a total of 210,336 data points.

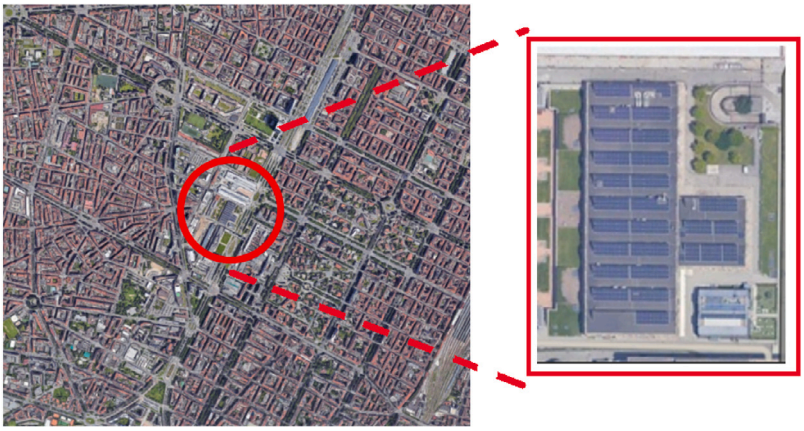

Fig. 1. University campus location in Turin, Italy, and its PV system under analysis.

The artifical and real datasets will then be further divided in different parts, each of which will be used in a different phase of the training, testing, exploitation and tuning of the ANN models and of the TL.

The simulated data is split and used to train and test the different ANN models. Years 2010 to 2014 (175,296 data points) are used as training set, while year 2015 (35,040 data points) is used as testing set. The real data set is then used to exploit the ANN models and evaluate their prediction performance against real data, and also to apply different TL techniques to tune the ANN models and to improve their prediction performance of PV power generation, always against the same real data. Year 2020 (35,136) is used to exploit the ANN models and evaluate their prediction performance against real data (inference set). Years 2018 and 2019 (70,080 data points) are used as training set to tune the ANN models with TL (tuning set), while year 2020 (inference set) is also used as test set for the TL models, and their performance is compared to that of the original ANN models exploited on the same year.

3. Methodology

This section aims to present the proposed methodology to forecast PV power generation when only a limited amount of real data is available, leveraging different meteorological data, a PV power generation simulator which accurately models a real PV installation, different ANNs and TL techniques.

As shown in Fig. 2, the different meteorological data features are first collected and then analyzed through feature selection methodologies in order to identify those which most impact the accuracy of data prediction forecast. The PV power generation simulator presented by [26], which is accurately modeled to replicate the real PV installation located on the roof of our university campus in Turin, Italy, is then used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. This artificial, but accurate and realistic dataset, together with the meteorological features previously selected, is used for the initial training and testing of the two different ANNs: 1D-CNN and LSTM. The resulting ANN models, trained and tested on the simulated dataset, are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data. Finally, different TL techniques are used to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data.

3.1. Data collection, preprocessing and feature selection

For the indicated time period, a dataset containing different meteorological features must be collected, together with the real PV generation data. The data must then be preprocessed, in order to make it coherent, for the whole dataset, and self-consistent. A selection of which of these features are most correlated to PV power generation is then carried out (Feature Selection), before feeding these features as inputs to the neural network models.

3.1.1. Data collection and preprocessing

The initial raw dataset, which is presented in the following list, is composed by 22 features and 385,718 rows, referring to years 2010 to 2020. All features are collected with 15 min intervals.

- Real PV power generation, i.e. real power (W) generated from the PV installation (data available from 2018 to 2020);
- GHI, Global Horizontal Irradiance (W/m^2);
- Relative Humidity, ranging between 0 and 1;
- Air temperature (°C);
- Air temperature - day (°C);
- Wind speed (m/s);
- Wind direction;
- Atmospheric pressure (hPa);
- Ultraviolet Index (UV index);
- Temperature (°C);
- Apparent temperature (°C);
- Humidity, ranging between 0 and 1;
- Wind speed (m/s) (wind speed Dark Sky);
- Wind bearing, the direction that the wind is coming from in degrees, with true north at 0° and proceeding clockwise;
- Dew point, the point at which dew can form (°C) below the atmospheric temperature (it changes with respect to pressure and humidity);
- Precipitation intensity, the intensity of precipitation at the given time;
- Precipitation probability, the probability of precipitation occurs, between 0 and 1;
- Cloud cover, the percentage of sky occluded by clouds, between 0 and 1;
- Simulated PV power (W) (simulated power);
- Sunset time, the unix timestamp when sun will set during a given day (s);
- Sunrise time, the unix timestamp when sun will rise during a given day (s);
- Sunshine duration, the difference between sunrise and sunset time (calculated);

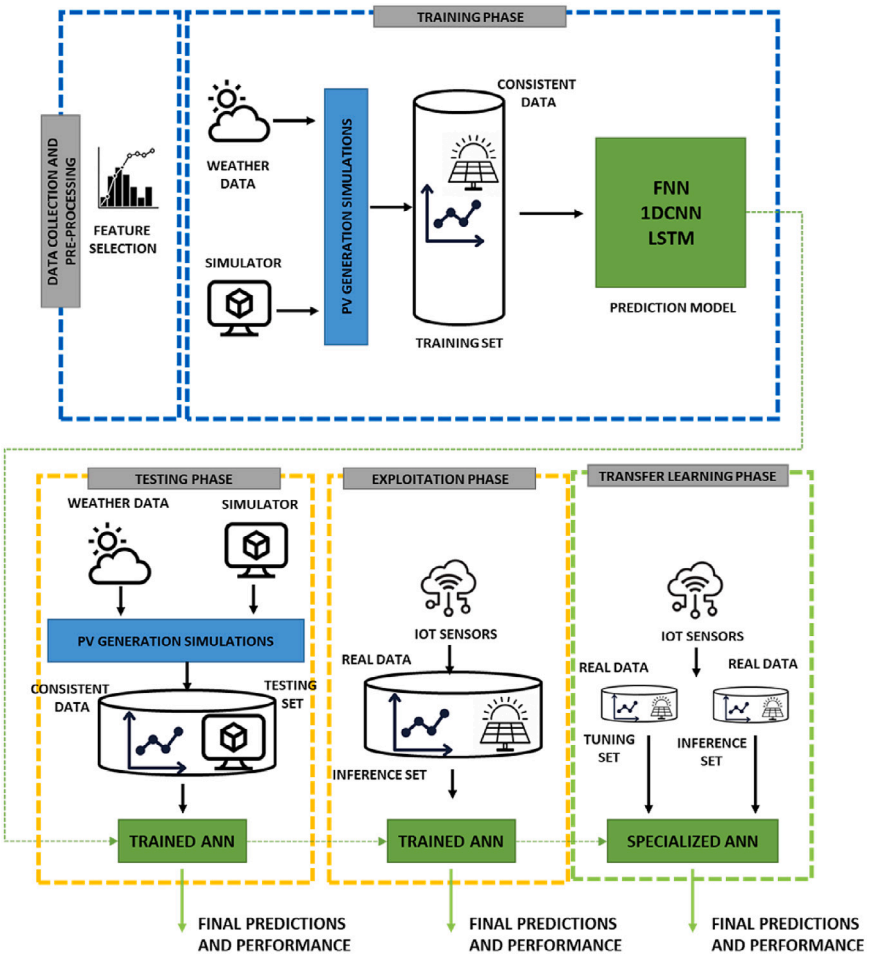

Fig. 2. Scheme of the proposed methodology.

- Day;
- Hour;
- Minute.

Real PV power generation is the real power (W) generated from the PV installation (data available from 2018 to 2020), and is provided by the sensors present in the PV installation in our university campus. The meteorological features are obtained from different sources. Features from GHI to atmospheric pressure are provided a weather station in our university campus, which is very close to the case study's building (see Section 2). Features from ultraviolet (UV) index to cloud cover are provided from Dark Sky [27], a software company specialized in weather observations and visualization which provides historical weather data in all the world, obtained from multiple sources (in Turin, the nearest station for data collection is located near the city's airport, 14.5 km north of the case study's building). The simulated power is obtain through the simulator [26], sunset and sunrise times are obtained through the pvlib python library [28], sunshine duration is calculated subtracting sunrise time from subset time, while day, hour and minute are embedded in the data.

After collecting the meteorological features' data, data cleaning and feature engineering steps were then carried out in order to clean the dataset.

Data cleaning

Table 2 presents a summary of the different features present in the dataset, highlighting the missing datapoints. The minimum and maximum values are also presented, after abnormal values for minimum and maximum were identified and replaced through linear interpolation.

In the period between the 6th of September 2016 until the 31st of December 2017, the weather station in our campus was not functioning and the data collected was therefore irregular or absent. The features affected by this issue are GHI, speed wind, atmospheric pressure, and wind direction. As a consequence, the data for that period is discarded (for a total number of 21,032 values), not just for the compromised features but, to be coherent, for the whole dataset, to make it self-consistent, and will not be used subsequently in any part of the training and/or validation of the neural networks and TL.

All features related to PV power generation were then set to 0 during the night (Real PV power generation, GHI, UV index, simulated power). This was done by leveraging the sunrise and sunset times. Specific missing values were then identified and replaced where possible. A strong, positive correlation between UV index and GHI was identified, so missing values in UV index were included based on existing UV index values for similar GHI. Finally, sporadic missing values in different features were replaced through linear interpolation.

Feature engineering

Feature engineering is a process used to manipulating raw data into features that can be effectively used in machine learning. More specifically, features scaling is required when the variables present very different orders of magnitude, and must therefore be normalized before being fed into the models. Normalized features all have the same magnitude, which increased the speed with which the machine learning models are trained. The min–max normalization is applied, scaling each variable in the range between 0 and 1.

It important to underline that min–max normalization is performed only during feature selection, to identify those features which most

Table 2
Raw dataset summary information.

	Data type	Min	Max	# of missing samples	Source
Real PV power generation	float64	0.00	562.20	0	Campus PV installation sensors
GHI	float64	0.00	1139.20	21,032	
Relative humidity	float64	0.09	1.00	8670	
Air temperature	float64	−10.90	36.20	880	Campus
Air temperature - day	float64	−6.00	33.20	772	Weather
Wind speed	float64	0.00	12.40	41,798	Station
Wind direction	float64	0.00	360.00	0	
Atmospheric pressure	float64	800.20	1008.40	41,798	
UV index	float64	0.00	10.00	12,666	
Temperature	float64	−11.60	36.20	5791	
Apparent temperature	float64	−13.50	36.20	5791	
Humidity	float64	0.05	1.00	4390	
Wind speed Dark Sky	float64	0.00	13.59	39,191	Dark
Wind bearing	float64	0.00	359.00	43,731	Sky
Dew point	float64	−22.25	24.32	1890	
Precipitation intensity	float64	0.00	12.954	45,968	
Precipitation probability	float64	0.00	1.00	45,968	
Cloud cover	float64	0.00	1.00	48,325	
Simulated power	float64	0.00	9.32	0	PV simulator
Sunrise time	int64	1 262 329 669.00	1 609 398 464.00	0	Embedded
Sunset time	int64	1 262 361 479.00	1 609 430 242.00	0	With other
Sunshine duration	int64	31 532.00	56 258.00	0	data

impact the models' ability to forecast PV power generation, aiming to reduce the number of input variables and therefore also the computational effort. Then, once the features which must be considered as inputs for the selected models have been identified, the works moves to the training and testing phases of the ANNs, during which there is no normalization.

3.1.2. Feature selection

The aim of the feature selection is to identify those features which most impact the models' ability to forecast PV power generation, aiming to reduce the number of input variables and therefore also the computational effort. [29] and [30] identify three main categories for feature selection:

- filter method: filtering is done using the correlation matrix and is most commonly carried out using Pearson Correlation.
- wrapper methods: it requires a machine learning algorithm and uses its performance as evaluation criteria.
- embedded methods: an iterative process, which analyzes each iteration of the model training process and extracts the features which contribute the most to the training.

In the following sections these techniques applied to our dataset are presented in detail.

3.1.3. Feature selection - filter methods

With filter methods, the selection of the features is independent from any machine learning algorithms, with features being selected based on their scores in statistical tests (correlation criteria). As presented by [29] and [30], since it does not rely on a machine learning algorithm, this technique avoids the risk of overfitting, but the selected subset of features are not optimal and could contain redundant variables. Pearson's correlation criteria and mutual information are the two ranking criteria used in our methodology.

Correlation Criteria

As presented by [29] and [30], the Pearson's correlation coefficient is used as a measure to quantify linear dependence between two variables. Its values vary from −1 to 1, where 1 means maximum positive linear correlation, −1 maximum negative linear correlation and 0 means no correlation.

Fig. 3, presents a correlation matrix with the results of the correlation coefficients computed between each variable, excluding the simulated power (since it is equivalent to the real powe). The correlation matrix reveals a positive linear correlation of 85% between UV index and GHI features, of 50% between real PV generation and UV index, and of 40% between real PV generation and GHI. Strong correlation (80%) appears to exist between sunshine duration and temperature, and between dew point and temperature, but such correlations are not useful for this investigation. The correlation between real PV generation and all other features appears to be very low, as that between GHI and all other features. Out of all variables, it is interesting to see how "sunshine duration" has a low correlation with "real PV power generation". As explained in Section 3.1.1, this variable is calculated and represents the difference between sunrise time (the unix timestamp when sun will set during a given day) and sunset time (the unix timestamp when sun will rise during a given day). This variable therefore does not take into consideration the weather conditions, and the actual amount of sun present. For this reason, this variable does not have such a strong correlation with "real PV power generation". Similar results for this variable will be confirmed also by other feature selection methods.

Since Pearson's correlation can be positive or negative, a new coefficient is adopted, the coefficient of determination R^2, which is defined as the square of the Pearson's coefficient R. As presented by [29] and [30], R^2 provides a more interpretable measure in order to compare the linear correlation between variables, and is used to evaluate the correlation between the different features and the real PV power generation, as presented in Fig. 4. According to these results, UV index is by far the feature with the highest correlation to PV power generation, with its R^2 coefficient at 0.25, almost 60% greater than the second feature, GHI. The next eight features (in decreasing order, GHI, sunset time, sunrise time, temperature, apparent temperature, humidity, relative humidity, and air temperature) fall between an R^2 range of 0.15 and 0.06. The next seven features (wind speed Dark Sky, dew point, air temperature - day, sunshine duration, wind speed, atmospheric pressure, wind bearing) present an R^2 between of 0.03 and 0.01, and the last features (precipitation intensity, precipitation probability, cloud cover, wind direction, day, hour, minute) present an R^2 of 0 a slightly above.

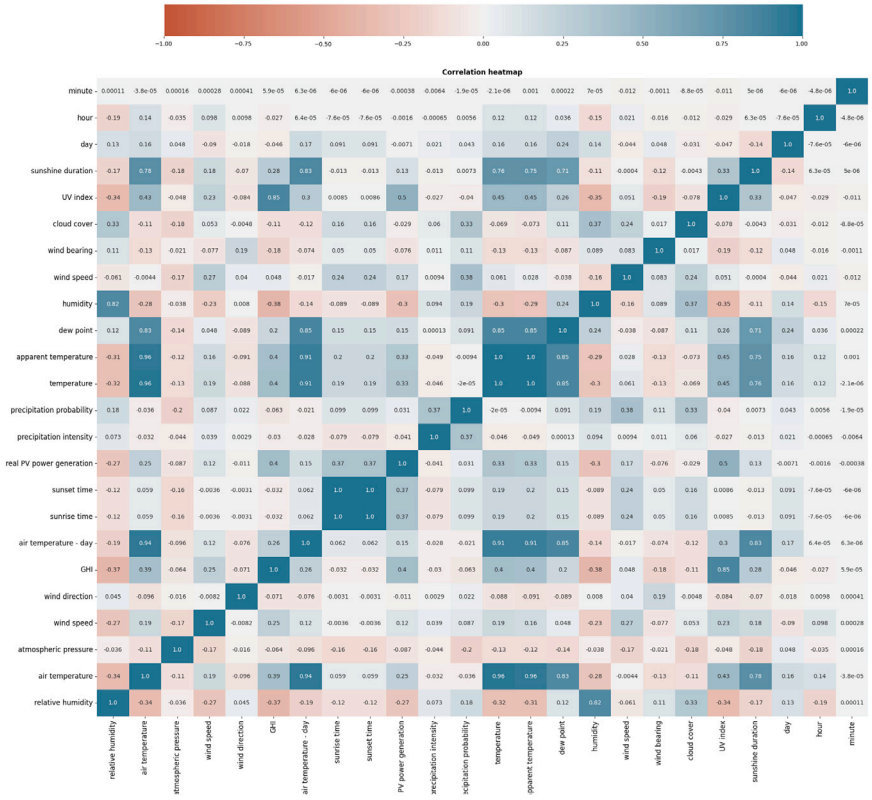

Fig. 3. Correlation matrix presenting correlation coefficients computed between each variable.

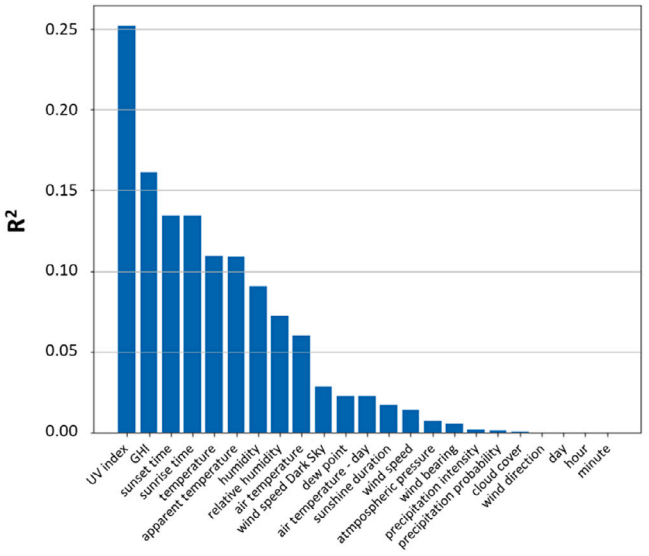

Fig. 4. Correlation criteria results between variables and P-PV Cittadella.

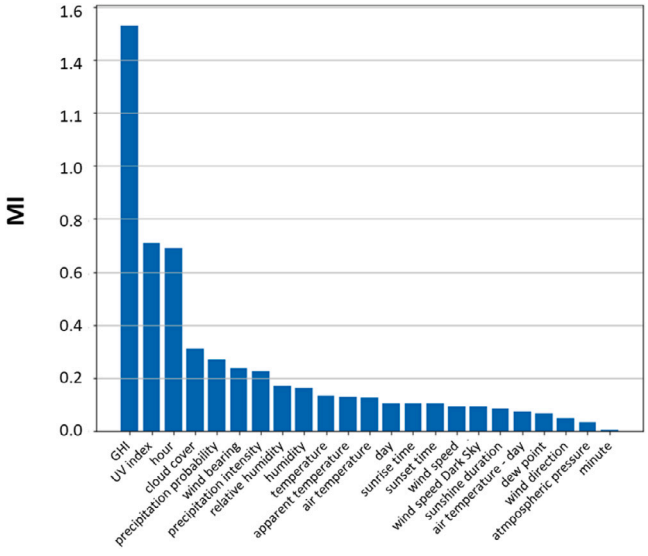

Fig. 5. Mutual information criteria results between variables and Real PV power generation.

Mutual information

The mutual information (MI) index measures how much can be learned from one variable by observing the other variable (or mutual dependence between two variables) [31], represented by the statistical dependence between the density of a variable x and the density of a variable y.

The probability densities of x and y are $p(x)$ and $p(y)$ respectively, with the joint probability density given by $p(x, y)$. Compared to the Pearson's correlation, the mutual information index is also able to recognize non linear correlations between variables. Fig. 5 presents the features sorted by their mutual information compared to the real PV power generation (P-PV Cittadella), with results showing how the best feature is by far the GHI, with a mutual information greater than 1.5, over twice that of the following features. UV index and hour are then respectively the second and third features, with mutual information around 0.7. The remaining twenty features have significantly lower mutual information, ranging between 0.3 and 0.

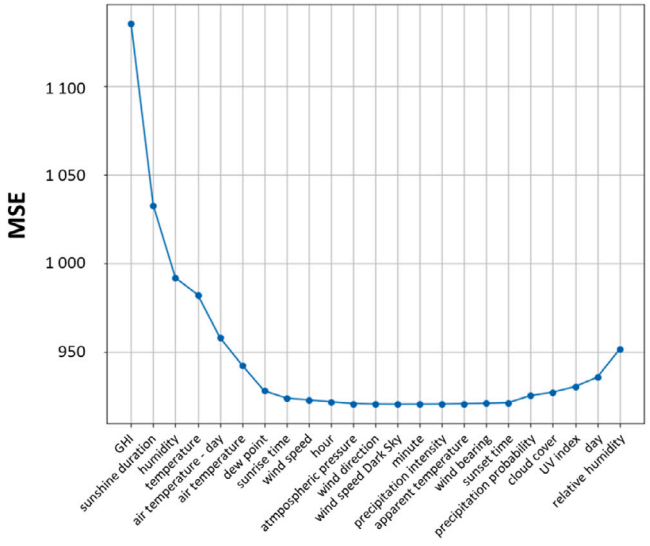

Fig. 6. SFS result criteria results between variables and Real PV power generation.

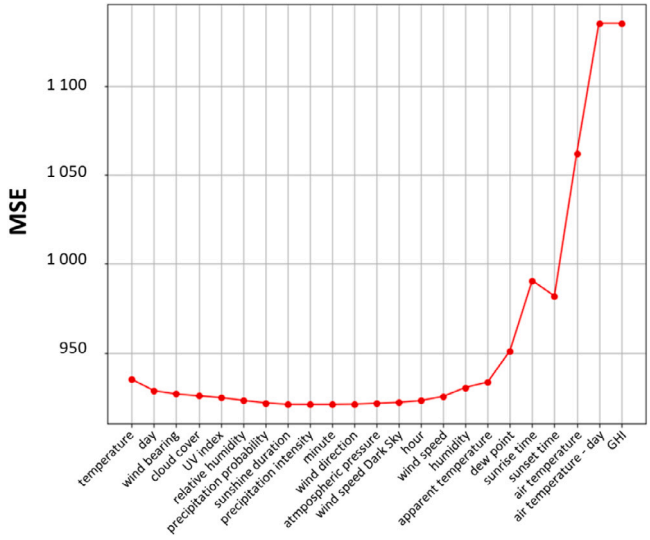

Fig. 7. SBS results between variables and Real PV power generation.

3.1.4. Feature selection - Wrapper method

In the Wrapper Method, the feature selection process is based on a specific machine learning algorithm, with the aim of solving an optimization problem by evaluating all the possible combinations of features, through an approach called greedy search, and finding the best ones compared to the real PV power generation. As presented by [29] and [30], this approach increases the performance and may reduce the risk of overfitting the data. However, as the number of features increases, the complexity also increases. The most commonly used algorithms are the sequential forward and the backward selection.

Sequential forward selection

The Sequential Forward Selection (SFS) algorithms stars with a null model, and at each step adds the features which give the highest performance. As presented by [29] and [30], this process continues until the addition of new features starts decreasing the performance. A disadvantage of this method is that it does not take into account dependencies among variables. In this work, a Linear Regression model is used. To train the model, the dataset needs to be split into training and testing. This division will be specific for the training of this model for this type of feature selection, and will therefore not follow the data division described in Section 2: this training set is composed of 2 years worth of data, from 2018 to 2019, while the test is performed on the last year of data (2020). After normalizing the features using the min–max normalization, the variables are fed to the algorithm and evaluated with respect to their Mean Squared Error (MSE) on the test set. Fig. 6 presents the results compared to the real PV power generation, with the most important features being GHI, sunshine duration and humidity, while the worst feature is the wind speed.

Sequential backward selection

While the sequential forward selection methods starts with an empty model, the Sequential Backward Selection (SBS) algorithm starts with a complete set of features and at each step removes those features which decrease the prediction performance. As presented by [29] and [30], the advantage of this method is that the features are evaluated in the presence of other variables, thus discarding useless features. The training and the testing set are the same of the sequential forward selection case. The algorithm deletes those features whose removal generated the lowest MSE on the test set. Fig. 7 shows the results compared to the real PV power generation, with most important features being GHI, air temperature and air temperature - day, while the worst is precipitation probability.

3.1.5. Feature selection - Embedded methods

In the embedded methods, the feature selection algorithm is integrated into the model's learning algorithm, it is incorporated in the training process and it combines the qualities of filter and of the wrapper methods. As presented by [29] and [30], embedded methods do not require to split the dataset into training and test sets, and the most common embedded technique are the LASSO regression and Random Forest.

LASSO Regression

Least Absolute Shrinkage and Selection Operator (LASSO) regression is a Linear Regression that uses l_1, and adds a regularization term called alpha to the cost function.

As presented by [29] and [30], LASSO regression is similar to features selection because l_1 regularization sets the weight of least significant features to zero. This regularization is done by varying the regularization of the hyperparameter $alpha$. The training set is composed by two years worth of data, from 2018 to 2020, with the features being normalized between 0 and 1 by using min–max normalization. The best model with lowest MSE value obtained from the training set is identifying by varying the model's hyperparameters, with $alpha$ equal to 0.001 being the best. Fig. 8 presents the features sorted by weight, with the best ones compared to the real PV power generation being GHI and temperature.

Random Forest

Random forest is a machine learning algorithm which combines the output of multiple decision trees to reach a single result. As presented by [29] and [30], each of these decision trees are trained on different random subsets of the training set. The most important features are located closer to the root of the tree, while the less important ones are close to the leaves. The drawback of this method is that the features could overfit the decision tree algorithm, discarding important features too. In this work, 50 estimators are used for a period ranging from 2018 to 2020. Fig. 9 presents the results compared to the real PV power generation, with GHI being by far the most important feature with a value greater than 0.8.

3.1.6. Results of features selection

Feature selection is used in order to find the most relevant features compared to the real PV power generation, which will then be used as inputs to the machine learning models, together with the PV power generation data. Table 3, presents the results for each feature selection

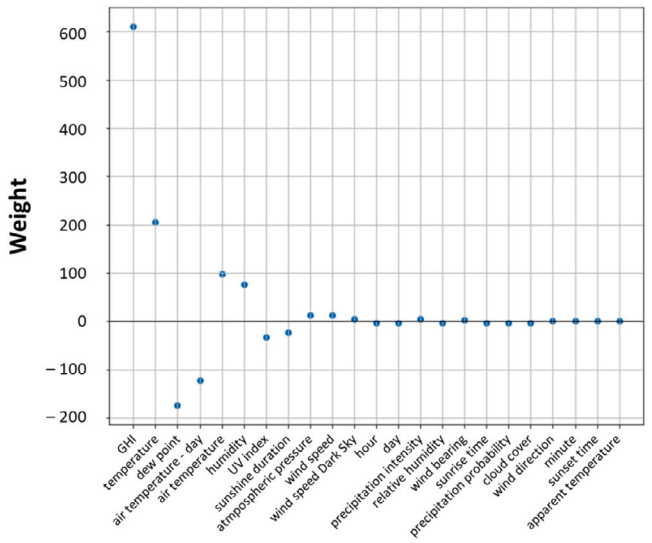

Fig. 8. LASSO Regression results between variables and Real PV power generation.

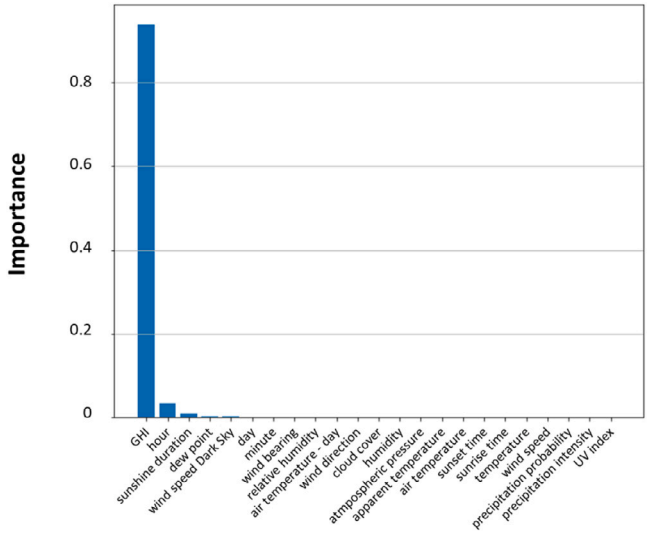

Fig. 9. Random Forest results between variables and Real PV power generation.

technique, with features sorted from most to least important. The ranking is obtained by calculating the mean of the results of each test, while the threshold between the selected and the discarded features was identified through a trial and error approach during the model evaluation.

The features selected for this investigations are GHI, humidity, temperature, dew point, UV index, sunshine duration and hour. It is interesting to notice how the different feature selection techniques present significant variance in classifying the features. Apart for GHI, which is consistently ranked as first or second for all feature selection techniques, the other features are evaluated very differently from the different methodologies. If one takes the top three features for each technique, they are not all part of the list of top features presented in Table 3. If one looks at the sequential backward selection technique, for example, only its first feature was included in the selected ones (GHI), while the others were all excluded. Similarly, for LASSO regression the second and third features rank as sixth and seventh overall, narrowly making the selected list. On the other hand, two to three of the final selected features rank over tenth in the different feature selection techniques.

Table 3
Feature selection results between variables and real PV power generation.

	Features	Methods						Rank
		R^2	MI	SFS	SBS	L	RF	
Selected	GHI	2	1	1	1	1	1	1
	Humidity	7	9	3	9	6	13	2
	Temperature	5	10	4	8	2	10	3
	Dew point	11	19	11	6	3	4	4
	UV index	1	2	9	16	7	23	5
	Sunshine duration	13	18	2	19	8	3	6
	Hour	22	3	15	11	12	2	7
Discarded	Air temperature	9	12	7	3	5	16	8
	Air temperature - day	12	20	5	2	4	10	9
	Relative humidity	8	8	6	17	15	9	10
	Sunrise time	4	15	13	4	17	18	11
	Day	21	14	8	10	13	6	12
	Wind speed Dark Sky	10	17	18	13	11	5	13
	Cloud cover	19	4	10	15	19	12	14
	Wind bearing	16	6	22	14	16	8	15
	Sunset time	3	13	23	5	22	17	16
	Apparent temperature	6	11	21	7	23	15	16
	Wind speed	14	16	14	12	10	20	17
	Precipitation probability	18	5	12	18	18	21	18
	Atmospheric pressure	15	22	16	20	9	14	19
	Precipitation intensity	17	7	20	23	14	22	20
	Wind direction	20	21	17	21	20	11	21
	Minute	23	23	19	22	21	7	23

3.2. Training, testing and exploitation of the neural networks

The purpose of this section is to present the methodology adopted during the development of the different ANN models. As presented in Fig. 2, the main phases required to develop any new predictive model are: (i) training (which includes training and validation), (ii) testing.

The PV power generation simulator presented by [26], which is accurately modeled to replicate the real PV installation located on the roof of our university campus in Turin, Italy, is used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. This artificial, but accurate and realistic dataset, together with the meteorological features previously selected in Section 3.1.6, is used for the initial training and testing of the two ANNs: 1D-CNN and LSTM. The resulting ANN models, trained and tested on the simulated dataset, are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data.

In this work, all the neural networks are trained using the Adaptive Moment Estimation (ADAM), with a learning rate of 0.001. Adaptive moment estimation is useful because it is able to avoid the cumbersome process of hyperparameters tuning dynamically the learning rate as needed, based on past gradient values. The loss function that has to be minimize during the training process is the MSE. Furthermore, in order to prevent overfitting and reduce the training time, the early stopping criteria with a patience of 10 epochs is introduced.

In order to identify the best architecture, the networks' performance is evaluated through four statistical indicators proposed by [32], being: Mean Absolute Difference (MAD), Mean Absolute Percentage Error (MAPE), Root Mean Square Difference (RMSD) and Coefficient of Determination (R^2).

The MAD, which measures the absolute difference between the prediction and the observed value:

$$MAD = \frac{\sum_{i=1}^{n} |y_{pred,i} - y_{test,i}|}{n} \quad (1)$$

The MAPE, which measures the relative difference between the prediction and the observed value:

$$MAPE = \frac{\sum_{i=1}^{n} \frac{|y_{pred,i} - y_{test,i}|}{y_{test,i}}}{n} 100 \quad (2)$$

The RMSD, which measures the standard deviation of the difference between predicted and observed values:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_{pred,i} - y_{test,i})^2}{n}} \tag{3}$$

The R^2, which measures the proportion of variance between the observed and the predicted values:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{test,i} - y_{pred,i})^2}{\sum_{i=1}^{n}(y_{test,i} - \bar{y}_{test})^2} \tag{4}$$

The ideal values of each metric used in this article, in terms of forecasting performance, were identified during past experiences, after reviewing the State of the Art literature on similar applications, particularly [21] and [30]. For example, the work of [21] presents R^2 = 0.8 as the threshold under which machine learning forecasting results cannot be considered acceptable. On the other hand, defining acceptable values for the other metrics, MAD, MAPE and RMSD, requires a more in-depth also of the data which wishes to be forecasted. This investigation focuses on effective forecasting of PV power generation which, according to Table 2, has values included between 0 and 562.20. If we consider 5% of the max PV power generation value, we have a value of approximately 28, which can be rounded up to 30. However, a difference of 30 between real and predicted value when the real value is close to its max, has an approximate error of 5%, but if the real value is close to 100, the approximate error results much higher. To properly identify the acceptable values for these metrics, we reviewed the approach presented by [21] and [30] on the same metrics, and we concluded that an acceptable metric for these models (LSTM and 1D-CNN) used to forecast this type of data (PV power generation) with this dataset, would be 20 for MAD, 50 for MAPE and 40 for RMSD.

3.2.1. Applicability of k-folds cross validation

When defining the investigation's methodology, existing literature was studied to identify different ways to ensure that the investigation's results are as robust as possible, and one of the methods which was found to be particularly well-established for machine learning predictive modeling is k-fold cross validation. However, after studying existing literature on the application of this method, such as [33,34] and [35], it was determined that k-folds cross validation is inapplicable to this work, due to the nature of the data which is the subject of the prediction.

The paper investigates innovative methods to predict photovoltaic power generation, by leveraging artificial neural networks which, in turn, leverage historical data for effective prediction. Photovoltaic power generation is strongly dependent on weather conditions, to the point that as presented by [21] and as applied in this investigation, meteorological features can be leveraged as an input to the artificial neural network models to successfully increase prediction accuracy for photovoltaic power generation.

However, due to the nature of the meteorological features data and, consequently, of the photovoltaic power generation data, coherence and continuity of the data is fundamental for the model to effectively learn from the data and successfully predict photovoltaic power generation. Macro-factors such as seasonality, and micro-factors such as changing weather conditions, strongly influence the varying photovoltaic power generation and must be captured by the model for it to accurately predict photovoltaic power generation. The application of k-folds cross validation, and of its random data sampling or random data splitting, would break the coherence and continuity of the data, and make it impossible for the model to successfully capture such trends.

More specifically, during the different seasons (i.e., winter, spring, summer, autumn) the trends of the meteorological features, in particular GHI, may seem similar (for example, during a sunny day without clouds the GHI values would always form a perfect-like bell curve), while in reality the absolute values are different: for example, in summer they are greater than in winter, as presented in Fig. 10. The application of k-folds cross validation, and of its random data sampling or random data splitting, would break the coherence and continuity of the data: if the Time Window (TW) are randomly split, there could be strong jumps between the last value of TW-1 and the first value of TW-2, and the model would be unable to effectively interpret the dataset and deliver accurate predictions.

A possible way to overcome this issue would be to divide the total dataset into four separate datasets based on the four seasons, and then divide each separate dataset into further subsets for training and testing, respectively. The investigation would therefore be training four distinct neural networks, one per season, and at this point the k-folds cross validation method could be applicable. However, the available dataset is not large enough to allow all these subdivisions: the four different training sets and four different test sets would not have sufficient data to train the models ([19] recommends selecting a data length of at least 3 years to effectively train artificial neural network models). This same problem is further accentuated for the real data dataset used when applying the different transfer learning techniques, dataset in which much less data is available.

A further challenge to the application of k-folds cross validation comes from occasional meteorological changes. For example, if at a certain point the sun is covered by temporary clouds, the photovoltaic power generation could see a temporary drop. However, if this event happens during the summer, in a relatively nice-weather week, the photovoltaic power generation in the hours and/or days before and after the cloud event would still be significant. Similarly, the meteorological features would also present specific behaviors which would be strongly influenced by both macro-factors such as seasonality, and micro-factors such as changing weather conditions. When the sun is covered by temporary clouds during the summer and in a relatively nice-weather week, certain meteorological features such as GHI and UV Index would see a significant drop, while others such as humidity and temperature would remain more stable. The model would therefore be able to capture these variations and locate them within a wider trend of data behavior, thus accurately leveraging the coherence and continuity of the data to accurately predict the photovoltaic power generation. The application of k-folds cross validation, and of its random data sampling or random data splitting, would break the coherence and continuity of the data, making it impossible for the model to successfully capture such trends, and therefore decreasing the accuracy of the photovoltaic power generation prediction.

Because of these reasons, it is clear that k-folds cross validation cannot be applied to this investigation.

3.2.2. 1D-CNN best architecture

The 1D-CNN is a particular kind of CNN, from which it differs from the dimension of the input and the way that the filter slides across the data. The CNN's filter (its ability to automatically detect important features), and its relative low cost makes it an extremely versatile model which is used in a variety of applications. Its application however is recommended when the dataset available for training is big enough in order to not occur in overfitting problem.

Also in this case a trial and error approach was used to identify the best architecture. The following hyperparameters were tested:

- number of one-dimensional convolution layers: varied between 1 and 2;
- number of units: varied between 10 and 200 (with an interval of 10 at a time);
- filter size: varied between 50 and 200 (first with an interval of 50 at a time, then between 150 and 200 with an interval of 10 at a time);
- kernel size: varied between 1, 2 and 3;
- activation functions: linear and hyperbolic, with and without a flatten layer and dense layer;

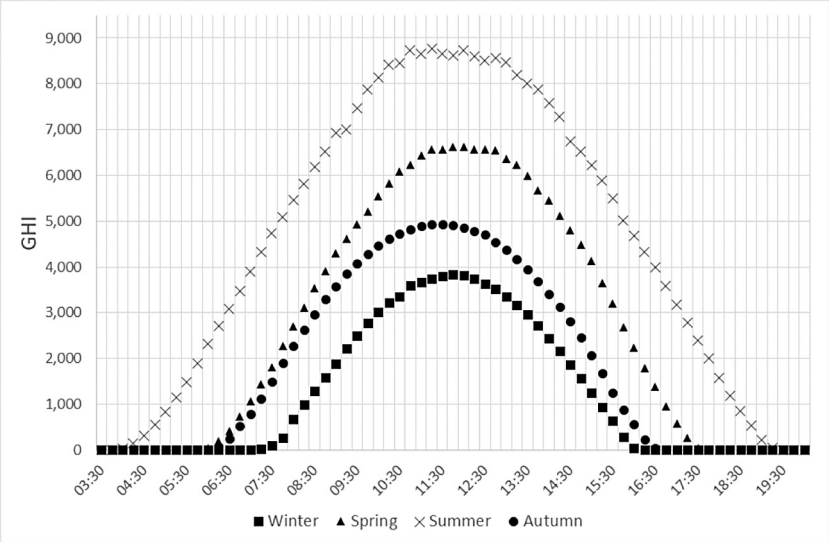

Fig. 10. Example of GHI values in a sunny day in winter, spring, summer and autumn.

- epochs: varied between 250 and 500 (with an interval of 50 at a time);
- batch size: varied between 100 and 400 (with an interval of 100 at a time).

The total number of possible combinations from this choice of ranges and intervals for the hyperparameters would have been over 20,000. As a consequence, a logical approach on how to reduce the overall combination was developed, based on our past experience working with these models and on the results presented in literature, for example by [21] and [30]. First, the number of layers was set at 2, and the activation functions were set at hyperbolic. Then, the kernel size was investigated at a high level, with parameters varying between 1, 2 and 3. Values of 2 and 3 clearly outperformed, and were chosen to continue. Then the filter size was investigated, with parameters varying between 50 and 200, with an interval of 50 at a time (50, 100, 15, 200). These different filter sizes were applied with both 1 and 2 kernel size. The results showed that kernel size 2 outperformed kernel size 3, and that filter sizes 150 and 200 outperformed the rest. A more capillary investigation of filter size was then carried out, varying between 150 and 200 with an interval of 10 at a time (150, 160, 170, 180, 190, 200), with kernel size 2. Filter size 170, with kernel size 2, outperformed the rest. These values were used to investigate the next parameter, batch size which varied between 100 and 400 with an interval of 100 at a time (100, 200, 300, 400). Batch size 200 outperformed the others. The number of epochs was then investigated, varying between 250 and 500 with an interval of 50 at a time (250, 300, 350, 400, 450, 500). 500 epochs outperformed the rest. Finally, the number of units was investigated, varying between 10 and 200, with an interval of 10 at a time (10, 20, 30, …, 180, 190, 200). 100 units appeared to be the ones with slightly better performance. Different activation functions were then investigated (hyperbolic and linear), together with different layers (with and without a flatten layer and dense layer). Also, the impact of a further fully connected layer, with varying units, added at the end of network using a hyperbolic tangent was also investigated. Furthermore, a pooling strategy is also used (Max Pooling) with pool size equal to 2, with the dimension of the output being halved as a consequence. Finally, the performance of the model with only 1 layer was also checked. Overall, 280 different combinations were tested.

3.2.3. LSTM best architecture

The LSTM neural network is a Recurrent Neural Network (RNN), useful for modeling sequential data. LSTM and RNNs also contain backward connections, meaning that at a given time t they receive the current state input x_t plus its own output at the previous time step y_{t-1}.

Also in this case a trial and error approach was used to identify the best architecture. The following hyperparameters were tested:

- number of layers: varied between 1, 2 and 3;
- number of units: varied between 10 and 150 (also varying between layers: with an interval of 10 at a time, and then 1 at a time between 20 and 30);
- activation functions: linear and hyperbolic (also varying between layers);
- epochs: varied between 100 and 500 (with an interval of 100 at a time);
- batch size: varied between 100 and 400 (with an interval of 100 at a time).

The total number of possible combinations from this choice of ranges and intervals for the hyperparameters would have been over 1200. As a consequence, as for the 1D-CNN a logical approach on how to reduce the overall combination was developed, based on our past experience working with these models and on the results presented in literature, for example by [21] and [30]. First, the number of layers was investigated, with parameters varying between 1, 2 and 3. Values of 2 and 3 clearly outperformed, and were chosen to continue. Then, for these two numbers of layers, the batch size was investigated, varying between 100 and 400 with an interval of 100 at a time (100, 200, 300, 400). Batch size 200, with 3 layers, outperformed the others. The number of epochs was then investigated, varying between 250 and 500 with an interval of 50 at a time (250, 300, 350, 400, 450, 500). 500 epochs outperformed the rest. Finally, the number of units was investigated, varying between 10 and 150, with an interval of 10 at a time (10, 20, 30, …, 130, 140, 150). 100 units appeared to be the ones with slightly better performance. Different activation functions were then investigated (hyperbolic and linear), and different tests were performed to see if changing some of these parameters within layers further influenced the performance. It was noticed that changes to the third layer had greatest impact. Overall, 130 different combinations were tested.

3.3. Transfer learning

The final section of the methodology presented in Fig. 2 aims to evaluate the effectiveness of TL in supporting PV power generation

forecast. Different TL techniques are used to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. As explained in Section 2, the data used for TL is now only the real, limited, dataset: years 2018 and 2019 (70,080 data points) are used for training (tuning), and year 2020 (35,136 data points) is used for testing.

Following the methodologies identified in the state of the art, TL can be applied through three main approaches:

- Retrain only the first layer
- Retrain only the second layer
- Retrain all the layers

As discussed in [36], and in [37], the different TL retraining approaches impact mainly accuracy and computation time. If all the layers of the model are retrained (no layers are frozen), the accuracy of the trained model will be greater, but will also require greater computation time. On the other hand, if only the first or the second layer are retrained (all layers but one are frozen), the model required backpropagating and updating the weights only of the unfrozen layer, which brings a significant decrease in computation time. The different solutions are therefore investigated.

For the purpose of this investigation, the performance of the different TL approaches was evaluated by considering only their accuracy (through the MAD, MAPE, RMSD and R^2 indexes). The computational time required for the three different TL approaches to complete the fine tuning of the neural networks was very similar, and in the order of minutes. The computational time required to train the original neural networks, on the other hand, was slightly longer, but still less than an hour, and once again similar for the two different models. Since the computational time required to train and fine tune the models was similar, it was not taken into consideration when evaluating their performance.

4. Results

In this section we report and discuss, for each ANN: the testing prediction performance of the most effective architectures in PV generation forecasting after their training with the artificial, but accurate and realistic, dataset of PV power generation created by the PV power generation simulator presented by [26]; the prediction performance of the same ANN models on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data; and the effectiveness of different TL techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, to improve the prediction performance of PV power generation always against the same real data.

For 1D-CNN, after training and testing the model with the artificial, but accurate and realistic, dataset of PV power generation created by the PV power generation simulator presented by [26], the best identified architecture consists of two 1-dimensional convolution layers with filter size equal to 170, kernel size equal to 2, a hyperbolic tangent as activation function, followed by a flatten and dense layer. Finally, a fully connected layer with 100 units is added at the end of the network, using a hyperbolic tangent. The output layer of the 1D-CNN was made up of 16 outputs, the number of epochs equals to 500 and the batch size is equal 200. Figs. 11–14 present the performance in terms of MAD, MAPE, RMSD and R^2 for the most effective 1D-CNN architectures. One can see that the variation between the different types of 1D-CNN architectures is limited, and all three indicators improve for the first four steps (up to 1 h), and then their performance deteriorates and appears to level out after 24 steps (6 h). The overall 1D-CNN performances are lower than the ones of the LSTM, both in terms of reduced variation and absolute value. The R^2, however, has similar values for both ANNs: at the 4th step (1 h) it is approximately 0.98; at the 8th step (2 h) it is approximately 0.88; and at the 12th step (3 h) it is approximately 0.85.

For LSTM, after training and testing the model with the artificial, but accurate and realistic, dataset of PV power generation created by the PV power generation simulator presented by [26], the best identified architecture consists of three recurrent layers, where the first two are composed by 100 units, with a hyperbolic activation function, and a third layer composed of 24 units that have a tanh activation function, and the output layer composed by 100 units with a linear activation function. Figs. 15–18 present the performance in terms of MAD, MAPE, RMSD and R^2 for the most effective LSTM architectures. One can see that the variation between the different types of LSTM architectures is limited, and all three indicators improve for the first four steps (up to 1 h), and then their performance deteriorates and appears to level out after 24 steps (6 h). The overall LSTM performances are overall better than the 1D-CNN ones. The RMSD, for example, at the 4th step (1 h) is approximately between 21 and 23 for 1D-CNN, and 20 for LSTM; at the 8th step (2 h) it is approximately 53 for 1D-CNN, 54 for LSTM; at the 12th step (3 h) it is approximately 60 for both ANNs.

Table 4 compares the prediction performance of the testing phase for the best of each of these ANNs (in terms of MAD, MAPE, RMSD and R^2), for different prediction horizons.

As far as MAD is concerned, one can see that the comparative performance between the different ANNs varies depending on the prediction horizon. In the first 30 min, 1D-CNN performance is slightly better than LSTM. From 15 to 45 min, performance of both LSTM and 1D-CNN actually improves, at 45 min they both perform at their best and LSTM start to slightly outperform 1D-CNN. From 60 min on, the performance of both networks worsens, but without major differences between them. For the MAPE results, the LSTM performance is always better than that of the 1D-CNN. For both models the MAPE also improves from 15 to 45 min, although with a countertrend at 30 min. Then from 60 min on, the performance of both networks worsens, with LSTM slightly outperforming 1D-CNN. For the RMSD results, LSTM performance is systematically slightly better than the 1D-CNN one and 1D-CNN networks. Again, from 15 to 45 min, performance of both LSTM and 1D-CNN actually improves, reaches its best at 45 min, and then deteriorates again from 60 min on. Finally, the R^2 results, show a much more homogeneous worsen in performance for both ANNs, with LSTM systematically overperforming the other, but by a matter of decimals. Again, from 15 to 45 min, performance of both LSTM and 1D-CNN actually improves, reaches its best at 45 min, and then deteriorates again from 60 min on.

The results clearly show that the best prediction performance is obtain with the LSTM model. The 1D-CNN also gives good prediction performance, and 1D-CNN also has better MAD values at the beginning with respect LSTM model, but the LSTM model generally outperforms the 1D-CNN one in most forecasting horizons for MAD, MAPE, RMSD, and R^2.

As explained in Section 3.2, and following the methodology presented in Fig. 2, the resulting ANN models, trained and tested on the simulated dataset, are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data. The previously presented best architectures for the ANN models are exploited on the inference set from the real dataset. Table 5 presents the prediction performance results on the real dataset for the 1D-CNN model, while Table 6 presents the results for the LSTM model. Results clearly show how the performance of both models deteriorate once they are exploited on a real dataset. The MAD, for example, deteriorates between 40% and 47% for 1D-CNN, and between 42% and 46% for LSTM. The MAPE also deteriorates, between 49% and 70% for 1D-CNN, and between 58% and 65% for LSTM. The RMSD deteriorates significantly, between 70% and 83% for 1D-CNN, and

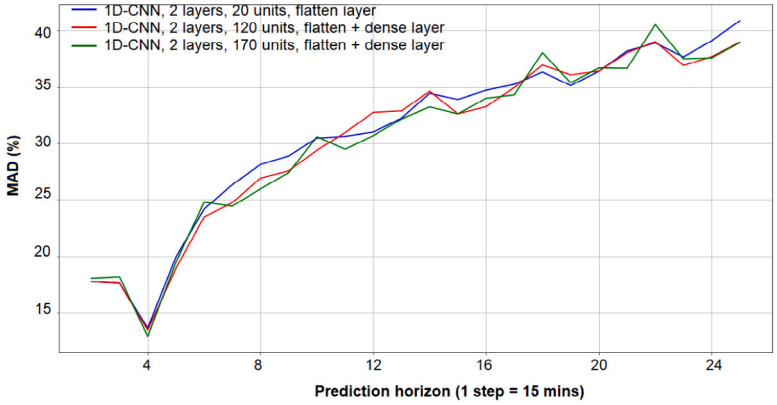

Fig. 11. Best 1D-CNN architecture MAD.

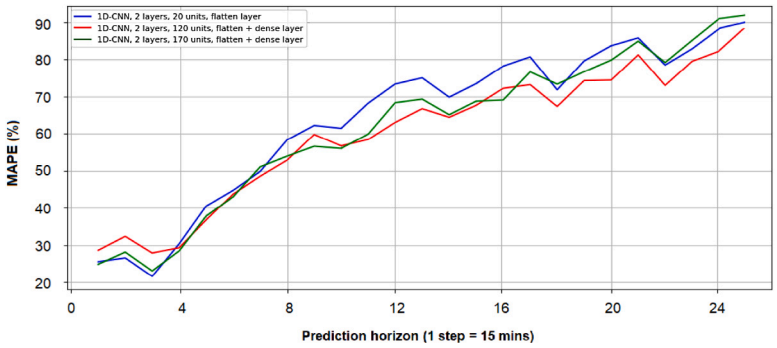

Fig. 12. Best 1D-CNN architecture MAPE.

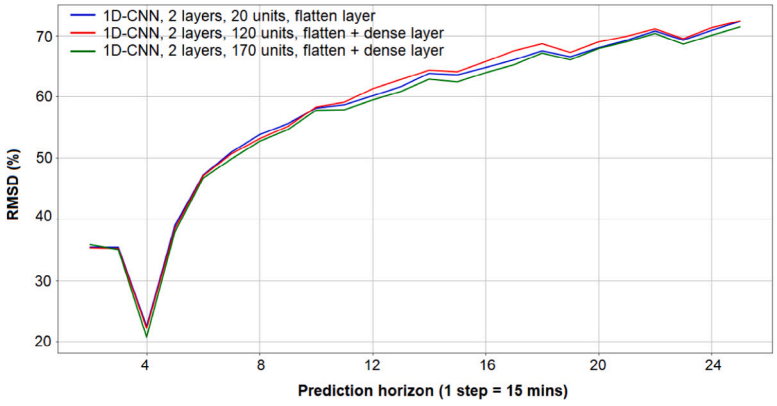

Fig. 13. Best 1D-CNN architecture RMSD.

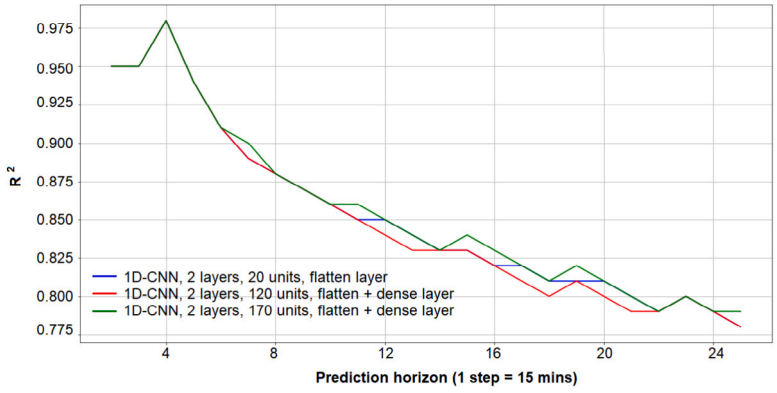

Fig. 14. Best 1D-CNN architecture R^2.

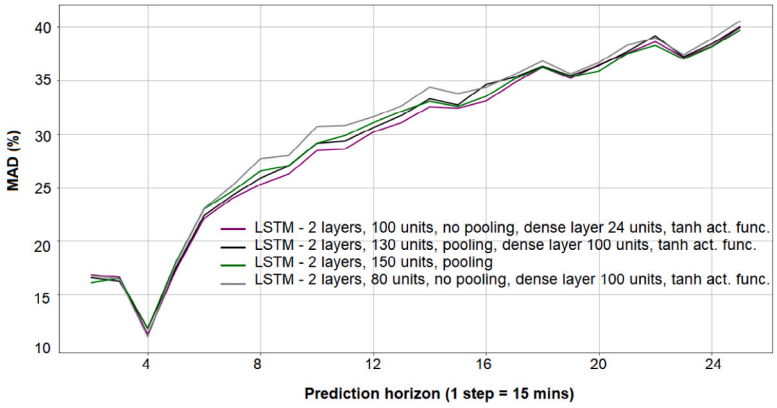

Fig. 15. Best LSTM architecture MAD.

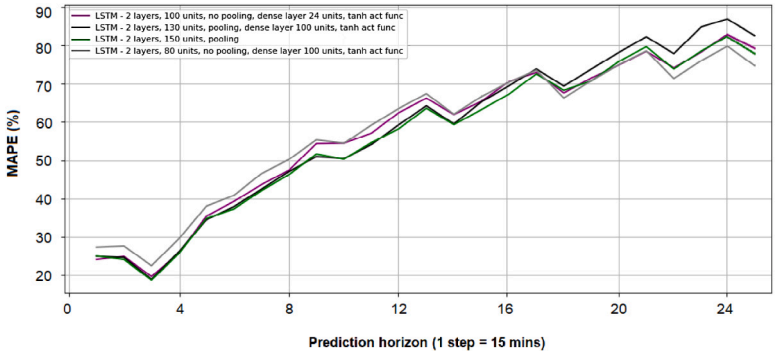

Fig. 16. Best LSTM architecture MAPE.

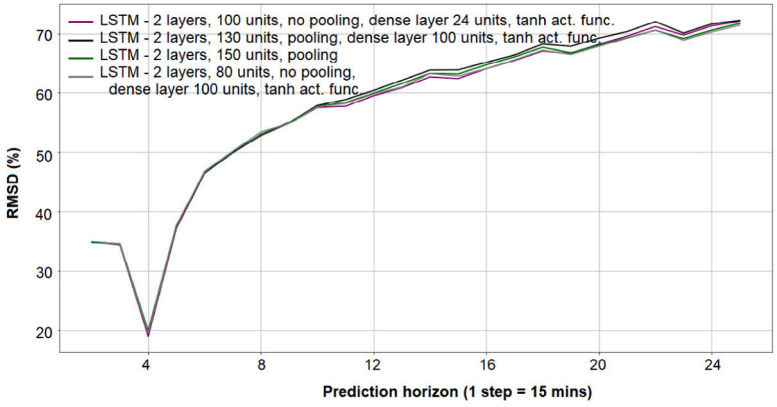

Fig. 17. Best LSTM architecture RMSD.

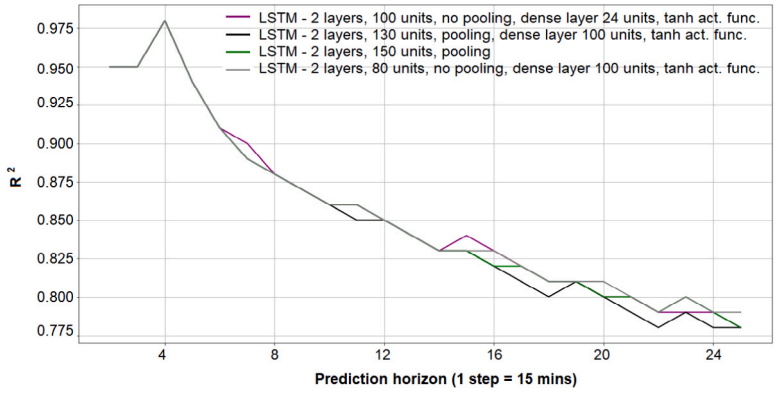

Fig. 18. Best LSTM architecture R^2.

Table 4
MAD, MAPE, RMSD and R^2 comparison for the best ANN architectures.

Pred. horizon	MAD		MAPE		RMSD		R^2	
	1D-CNN	LSTM	1D-CNN	LSTM	1D-CNN	LSTM	1D-CNN	LSTM
15 min	**19.27**	19.63	25.37	**24.43**	35.70	**35.38**	0.95	**0.95**
30 min	**17.87**	17.95	28.12	**25.12**	35.08	**34.69**	0.95	**0.95**
45 min	12.87	**12.80**	23.72	**19.76**	20.94	**20.42**	0.98	**0.98**
60 min	19.26	**18.86**	29.15	**27.43**	38.03	**37.69**	0.94	**0.94**
75 min	24.96	**23.41**	38.64	**36.62**	46.95	**46.71**	0.91	**0.91**
90 min	24.89	**24.30**	44.32	**39.43**	50.28	**50.01**	0.89	**0.89**
105 min	27.22	**26.10**	52.86	**43.89**	53.73	**52.92**	0.88	**0.88**
120 min	28.33	**27.61**	54.57	**47.68**	55.61	**54.78**	0.87	**0.87**

Table 5
1D-CNN MAD, MAPE, RMSD, and R^2 comparison between testing and exploitation prediction performance.

Pred. horizon	MAD		MAPE		RMSD		R^2	
	Testing	Exploitation	Testing	Exploitation	Testing	Exploitation	Testing	Exploitation
15 min	19.27	27.36	25.37	37.67	35.70	62.48	0.95	0.80
30 min	17.87	25.02	28.12	46.40	35.08	59.64	0.95	0.81
45 min	12.87	18.28	23.72	40.32	20.94	37.27	0.98	0.78
60 min	19.26	27.16	29.15	48.24	38.03	67.69	0.94	0.78
75 min	24.96	35.44	38.64	64.92	46.95	83.57	0.91	0.78
90 min	24.89	36.09	44.32	75.12	50.28	90.50	0.89	0.77
105 min	27.22	40.01	52.86	89.33	53.73	98.33	0.88	0.76
120 min	28.33	41.65	54.57	91.13	55.61	101.77	0.87	0.75

Table 6
LSTM MAD, MAPE, RMSD and R^2 comparison between testing and exploitation prediction performance.

Pred. horizon	MAD		MAPE		RMSD		R^2	
	Testing	Exploitation	Testing	Exploitation	Testing	Exploitation	Testing	Exploitation
15 min	19.63	28.07	24.43	40.19	35.38	62.27	0.95	0.79
30 min	17.95	25.49	25.12	40.44	34.69	60.36	0.95	0.80
45 min	12.80	18.18	19.76	31.22	20.42	36.76	0.98	0.77
60 min	18.86	26.97	27.43	43.89	37.69	68.22	0.94	0.76
75 min	23.41	33.48	36.62	57.86	46.71	84.55	0.91	0.76
90 min	24.30	34.75	39.43	65.26	50.01	91.02	0.89	0.75
105 min	26.10	37.58	43.89	71.76	52.92	96.84	0.88	0.74
120 min	27.61	40.31	47.68	75.57	54.78	101.34	0.87	0.73

between 74% and 85% for LSTM. Finally, the R^2 deteriorates between 15% and 20% for 1D-CNN, and between 16% and 21% for LSTM but, most importantly, it immediately deteriorates below 0.8.

As explained in Section 3.2, and following the methodology presented in Fig. 2, different TL techniques are then used to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. The tuning set of the real dataset is therefore used to re-train the ANN models with the TL methodology, while the inference set of the real dataset is used for testing, and comparing the prediction performance on real data of TL to that of the exploited models.

As presented in Section 3.3, different TL techniques are used: one where only the first layer is retrained, one where only the last layer is retrained, and one where all layers are retrained. For each ANN, the prediction performance of the original models presented in the exploitation phase, without TL, and the prediction performance of the three different TL techniques, is presented in the following figures. For the 1D-CNN, Figs. 19–22 present the MAD, MAPE, RMSD and R^2 performance of the different TL techniques. Figs. 23–26, on the other hand, present the MAD, MAPE, RMSD and R^2 performance of the different TL techniques applied to the LSTM.

For 1D-CNN, the best performance is obtained with the second TL model (where only the second layer is retrained) which, however, delivers only marginal performance improvement: up to 30 min for MAD and RMSD, and up to 75 min for R^2, while for MAPE this TL model outperforms for up to 2 h and then converges. For LSTM on the other hand, it is the third TL model (where all layers are retrained), which outperforms the others, and which is able to deliver significant performance improvement: up to 90 min for MAD and RMSD, up to 120 min for R^2, while for MAPE it underperforms for the first 30 min and remains more performing for up to 2 h.

The performance of the 1D-CNN model presented in the test, exploitation and best TL phases, are compared in Table 7. Similarly, for LSTM the achieved results are presented by Table 8.

The difference in variation of performance of the 1D-CNN and LSTM models from exploitation to TL are summarized in Table 9. A positive variation indicates an improvement in the application of TL, while negative variation indicates a worse performance of TL.

When applying TL to the 1D-CNN, at the first 15 min forecast the MAD improves by 12.0%, RMSD by 17.1% and R^2 by 21.9%, while the MAPE is slightly worst by 5.5%. At the 30 min forecast, the MAD improvement decreases to 3.6%, RMSD to 2.9% and R^2 to 9.1%, while the MAPE improves to 10.9%. After 30 min however, the MAD and RMSD show a worsening in performance, while the R^2 continues to remain better until 75 min. After 90 min, all three indicators underperform with TL. The MAPE instead slightly underperforms again at 45 min to −3.1%, but then improves continuously until 105 min, and at 2 h its improvements appears to start decreasing.

For the LSTM on the other hand, TL appears to be even more effective. For the 15 min forecast the MAD improves by 15.4%, MAPE by 39.5%, RMSD by 18.2% and R^2 by 20.6%. For the 30 min forecast the MAD remains at 9.9%, MAPE by 26.5%, RMSD at 6.9% and R^2 at 14.1%. At 45 min, MAPE slightly worsens to −7.2%, but then improves again immediately, and performance remains better for all indicators until 90 min. After 90 min, the MAD and RMSD appear to converge, with differences in performance between exploitation and TL leveling out around 0. R^2 with TL, on the other hand, continues to overperfom

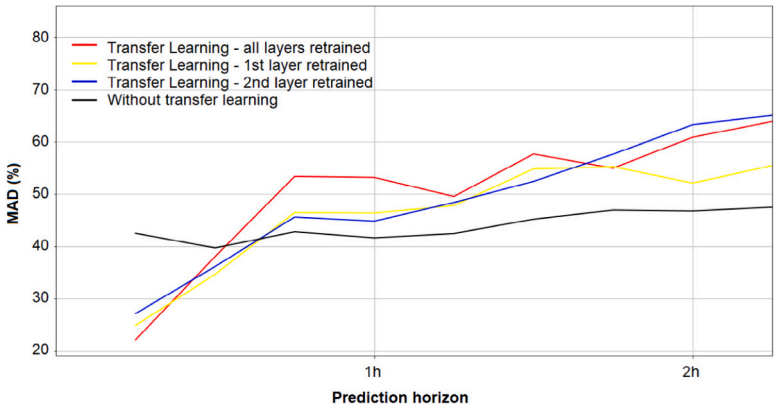

Fig. 19. 1D-CNN MAD comparison with TL.

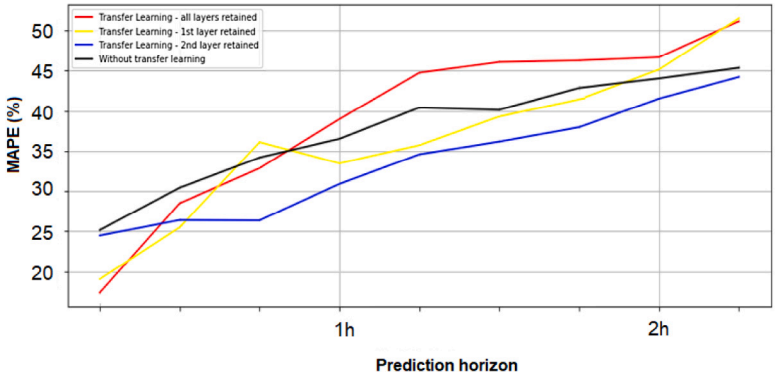

Fig. 20. 1D-CNN MAPE comparison with TL.

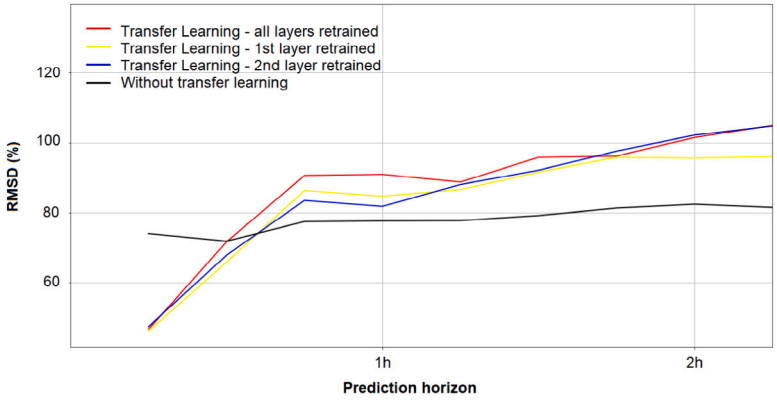

Fig. 21. 1D-CNN RMSD comparison with TL.

up to 2 h, with stable values of around 10% better than without TL. Similarly, also MAPE continues to overperfom up to 2 h, with performance actually improving up to 44% better than without TL.

The above results therefore clearly show that TL brings limited improvement to the 1D-CNN, but consistent improvement to LSTM, which remains the best performing ANN even after TL.

5. Conclusion

The purpose of this investigation is to present an innovative methodology for PV power generation forecasting with ANNs, when only a limited amount of real data is available. The novelty of this investigation lies in the exploitation of a PV power generation simulator, which accurately models a real PV installation, to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANN models, which are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data. Further novelty is brought by the application of different TL techniques to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data.

A variety of features obtained from different meteorological data are first analyzed through feature selection methodologies in order to identify those which most impact the accuracy of data prediction forecast. The variables which most impact power prediction are found to be GHI, humidity, temperature, dew point, UV index, sunshine duration and

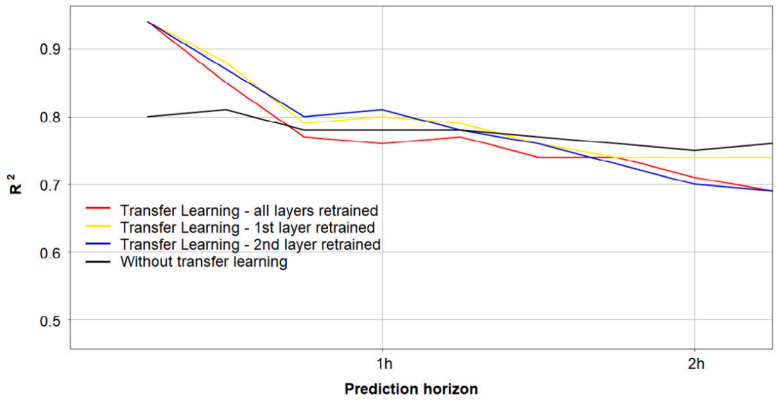

Fig. 22. 1D-CNN R^2 comparison with TL.

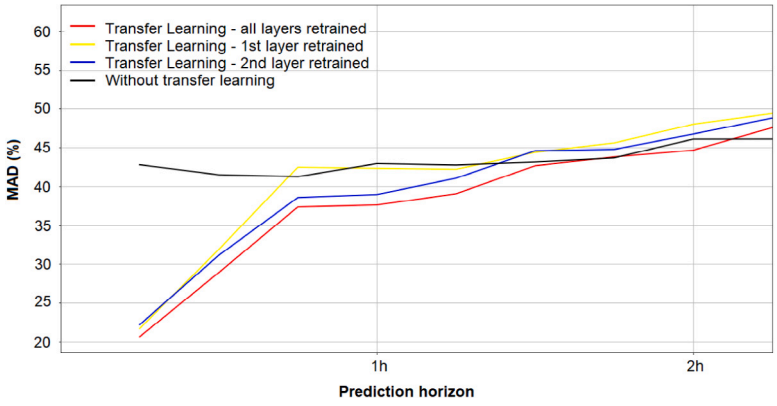

Fig. 23. LSTM MAD comparison with TL.

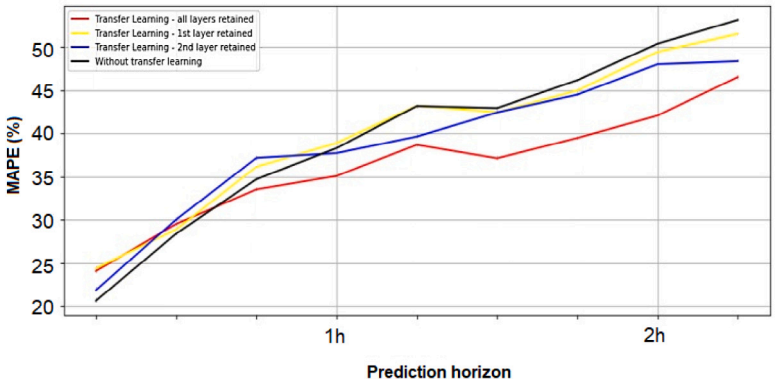

Fig. 24. LSTM MAPE comparison with TL.

time of the day. The PV power generation simulator presented by [26], which is accurately modeled to replicate real PV installations, is then used to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANNs. This artificial, but accurate and realistic dataset, together with the meteorological features previously selected, is used for the initial training and testing of the two different ANNs: 1D-CNN and LSTM.

The resulting ANN models, trained and tested on the simulated dataset, are then exploited on a portion of real, but limited, dataset of the real power generated by the real PV installation on which the simulator is modeled, to evaluate their prediction performance against real data. Finally, different TL techniques are used to tune the ANN models with the remaining portion of the real, but limited, dataset of PV power generation, investigating their effectiveness to improve the prediction performance of PV power generation always against the same real data. The methodology has been tested and validated on a real-life PV installation located on the rooftop of a building of our university campus in Turin, Italy.

For LSTM, the best identified architecture consists of three recurrent layers, where the first two are composed by 100 units, with a hyperbolic activation function, and a third layer composed of 24 units that have a tanh activation function, and the output layer composed by 100 units with a linear activation function. Furthermore, TL applied to this LSTM architecture allows improvement in performance, especially when applying the third TL model, where all layers are retrained. The improvement is significant in the short-term (up to 30 min), and slight up to 120 min.

For 1D-CNN, the best identified architecture consists of two 1-dimensional convolution layer with filter size equal to 170, kernel size

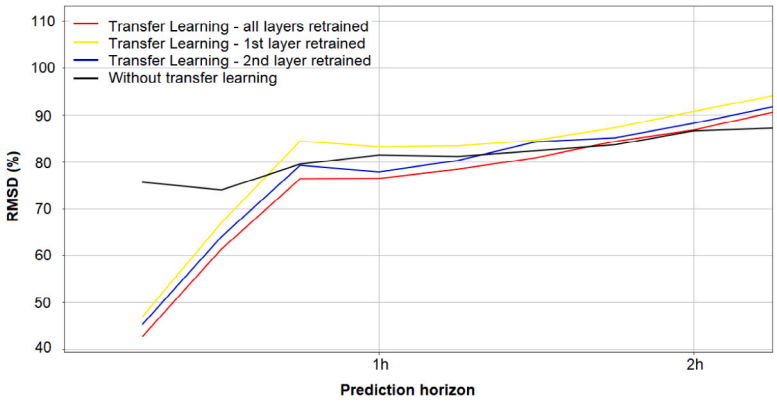

Fig. 25. LSTM RMSD comparison with TL.

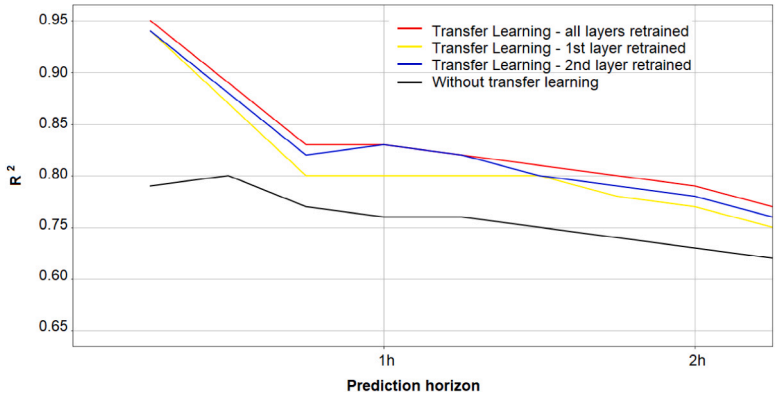

Fig. 26. LSTM R^2 comparison with TL.

Table 7
1D-CNN MAD, MAPE, RMSD and R^2 comparison between testing, exploitation and best TL prediction performance.

Pred.	MAD			MAPE		
horizon	Testing	Exploitation	TL	Testing	Exploitation	TL
15 min	19.27	27.36	24.09	25.37	37.67	39.75
30 min	17.87	25.02	24.12	28.12	46.40	41.32
45 min	12.87	18.28	18.79	23.72	40.32	41.58
60 min	19.26	27.16	28.12	29.15	48.24	46.61
75 min	24.96	35.44	37.19	38.64	64.92	49.83
90 min	24.89	36.09	38.58	44.32	75.12	51.13
105 min	27.22	40.01	42.19	52.86	89.33	53.21
120 min	28.33	41.65	43.06	54.57	91.13	56.83
Pred.	RMSD			R^2		
horizon	Testing	Exploitation	TL	Testing	Exploitation	TL
15 min	35.70	62.48	51.77	0.95	0.80	0.98
30 min	35.08	59.64	57.88	0.95	0.81	0.88
45 min	20.94	37.27	38.74	0.98	0.78	0.81
60 min	38.03	67.69	69.98	0.94	0.78	0.82
75 min	46.95	83.57	86.86	0.91	0.78	0.78
90 min	50.28	90.50	95.53	0.89	0.77	0.76
105 min	53.73	98.33	104.77	0.88	0.76	0.72
120 min	55.61	101.77	108.44	0.87	0.75	0.68

Table 8
LSTM MAD, MAPE, RMSD R^2 comparison between testing, exploitation and best TL prediction performance.

Pred.	MAD			MAPE		
horizon	Testing	Exploitation	TL	Testing	Exploitation	TL
15 min	19.63	28.07	23.75	24.43	40.19	24.32
30 min	17.95	25.49	22.98	25.12	40.00	29.91
45 min	12.80	18.18	17.54	19.76	31.22	33.47
60 min	18.86	26.97	25.84	27.43	43.89	35.12
75 min	23.41	33.48	32.54	36.62	57.86	38.22
90 min	24.30	34.75	34.51	39.43	65.26	37.84
105 min	26.10	37.58	37.58	43.89	71.76	39.73
120 min	27.61	40.31	40.03	47.68	75.57	42.12
Pred.	RMSD			R^2		
horizon	Testing	Exploitation	TL	Testing	Exploitation	TL
15 min	35.38	62.27	50.95	0.95	0.79	0.95
30 min	34.69	60.36	56.20	0.95	0.80	0.91
45 min	20.42	36.76	35.94	0.98	0.77	0.85
60 min	37.69	68.22	66.33	0.94	0.76	0.85
75 min	46.71	84.55	83.14	0.91	0.76	0.84
90 min	50.01	91.02	90.52	0.89	0.75	0.83
105 min	52.92	96.84	97.37	0.88	0.74	0.82
120 min	54.78	101.34	101.34	0.87	0.73	0.81

equal to 2, a hyperbolic tangent as activation function, followed by a flatten and dense layer. Finally, a fully connected layer with 100 units is added at the end of the network, using a hyperbolic tangent. The output layer was made up of 16 outputs, the number of epochs equals to 500 and the batch size is equal 200. TL applied to this 1D-CNN architecture allows some improvement in performance, when applying the second TL model, where only the second layer is retrained. This improvement though is only slight, and only in the short-term (up to 30 min).

Results therefore show how 1D-CNN and LSTM ANNs can successfully be used for PV power generation predictions even if trained on simulated datasets. Furthermore, the application of TL techniques has proved useful to improve the performance of PV power forecasting, both in the short-term (15-30 min, for 1D-CNN) and in the mid-term (up 2 h, for LSTM). These conclusions open some interesting reflections on the advantages and disadvantages coming from the proposed methodology.

Table 9
Variation in performance variation from exploitation to TL for 1D-CNN and LSTM.

Pred. horizon	1D-CNN variation				LSTM variation			
	MAD [%]	MAPE [%]	RMSD [%]	R^2 [%]	MAD [%]	MAD [%]	RMSD [%]	R^2 [%]
15 min	12.0	−5.5	17.1	21.9	15.4	39.5	18.2	20.6
30 min	3.6	10.9	2.9	9.1	9.9	26.5	6.9	14.1
45 min	−2.8	−3.1	−3.9	3.3	3.5	−7.2	2.2	10.1
60 min	−3.5	3.4	−3.4	4.9	4.2	20.0	2.8	12.1
75 min	−4.9	23.2	−3.9	0.0	2.8	33.9	1.7	10.4
90 min	−6.9	31.9	−5.6	−1.7	0.7	42.0	0.5	10.7
105 min	−5.4	40.4	−6.6	−5.2	0.0	44.6	−0.5	11.0
120 min	−3.4	37.6	−6.6	−8.9	0.7	44.3	0.0	11.3

An advantage provided by this investigation is the understanding that a PV power generation simulator, which accurately models a real PV installation, can be trusted to create an artificial, but accurate and realistic, dataset of PV power generation large enough to effectively train and test different ANN models, and that these models can then be exploited on a more limited, but real dataset of the real power generated by the real PV installation on which the simulator is modeled, to predict with acceptable accuracy the PV power generation of the PV installation. This is a very interesting advantage, since it allows to overcome the difficulty of obtaining large, reliable datasets of real data, which is required to train ANN models to an acceptable level of accuracy. Another minor advantage of this investigation is the understand that 1D-CNN and, especially, LSTM are two ANN models which can successfully be used for PV power generation predictions even if trained on simulated datasets. This advantage allows future works and applications to focus on ANN models which have been proved to perform positively, and also provide some insight on what parameters are successful with these models, thus reducing any time which would otherwise be dedicated to trial and error activities trying to find the most suitable models and their parameters. A further advantage provided by this investigation is the understanding that the application of TL techniques on these 1D-CNN and LSTM models can be useful to improve the performance of PV power forecasting, both in the short-term (15-30 min, for 1D-CNN) and in the mid-term (up 2 h, for LSTM), also with limited real data available. This advantage is quite important for two reasons: first of all, it can avoid future applications from going through the lengthy process of training new ANNs from scratch, thus saving potentially months of activity. Furthermore, in case future applications wish to apply ANNs for PV power forecasting of a specific PV installation, but do not have a large, reliable dataset of real data, which is required to train ANN models to an acceptable level of accuracy, if they have an ANN already trained on another PV installation, they can consider applying TL techniques on the existing ANN, and potentially receive good performance both in the short-term (15-30 min) and in the mid-term (up 2 h). This last advantage, however, is laid with uncertainty.

On the other hand, some disadvantages provided by this investigation is the understanding that the whole methodology was successful but that it relied on the exploitation of a PV power generation simulator, which accurately models a real PV installation. If such an accurate simulator is unavailable, it is unknown if similar positive results are still achievable. If, for example, future applications wish to leverage this methodology to adapt an ANN already trained on another, different, PV installation, TL could be successful but there is no actual data supporting this possibility. However, this uncertainty will be the subject of future works and investigations.

CRediT authorship contribution statement

Andrea Bellagarda: Methodology, Software, Data curation, Writing. **Donato Grassi:** Methodology, Software, Data curation. **Alessandro Aliberti:** Methodology, Supervision, Review & editing. **Lorenzo Bottaccioli:** Conceptualization, Validation. **Alberto Macii:** Conceptualization, Validation. **Edoardo Patti:** Methodology, Supervision, Review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] I.I.E. Agency, World Energy Outlook 2022, Tech. Rep., IEA - International Energy Agency, 2022, URL https://www.iea.org/reports/world-energy-outlook-2022.

[2] H. Chandler, Empowering Variable Renewables-Options for Flexible Electricity Systems, IEA, Paris, 2008.

[3] S. Impram, S. Varbak Nese, B. Oral, Challenges of renewable energy penetration on power system flexibility: A survey, Energy Strategy Rev. 31 (2020) 100539, http://dx.doi.org/10.1016/j.esr.2020.100539, URL https://www.sciencedirect.com/science/article/pii/S2211467X20300924.

[4] L.T. Berger, K. Iniewski, Smart Grid Applications, Communications, and Security, John Wiley & Sons, 2012.

[5] P. Siano, Demand response and smart grids—A survey, Renew. Sustain. Energy Rev. 30 (2014) 461–478.

[6] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F.M. de Pison, F. Antonanzas-Torres, Review of photovoltaic power forecasting, Sol. Energy 136 (2016) 78–111, http://dx.doi.org/10.1016/j.solener.2016.06.069, URL https://www.sciencedirect.com/science/article/pii/S0038092X1630250X.

[7] F. Wang, Z. Mi, S. Su, H. Zhao, Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters, Energies 5 (2012) 1355–1370, http://dx.doi.org/10.3390/en5051355.

[8] H. Wang, G. Wang, G. Li, J. Peng, Y. Liu, Deep belief network based deterministic and probabilistic wind speed forecasting approach, Appl. Energy 182 (2016) 80–93, http://dx.doi.org/10.1016/j.apenergy.2016.08.108, URL https://www.sciencedirect.com/science/article/pii/S0306261916312053.

[9] H.T. Pedro, C.F. Coimbra, Assessment of forecasting techniques for solar power production with no exogenous inputs, Sol. Energy 86 (7) (2012) 2017–2028, http://dx.doi.org/10.1016/j.solener.2012.04.004, URL https://www.sciencedirect.com/science/article/pii/S0038092X12001429.

[10] A. Nespoli, E. Ogliari, S. Leva, A. Massi Pavan, A. Mellit, V. Lughi, A. Dolara, Day-ahead photovoltaic forecasting: A comparison of the most effective techniques, Energies 12 (9) (2019) 1621.

[11] S.S. Haykin, Neural Networks: A Comprehensive Foundation, Macmillan, New York, 1994, Republished in 1998.

[12] M. Pierro, M. De Felice, E. Maggioni, D. Moser, A. Perotto, F. Spada, C. Cornaro, Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data, Sol. Energy 158 (2017) 1026–1038.

[13] C. Chen, S. Duan, T. Cai, B. Liu, Online 24-h solar power forecasting based on weather type classification using artificial neural network, Sol. Energy 85 (11) (2011) 2856–2870.

[14] M.G. De Giorgi, P.M. Congedo, M. Malvoni, Photovoltaic power forecasting using statistical methods: impact of weather data, IET Sci. Meas. Technol. 8 (3) (2014) 90–97.

[15] U.K. Das, K.S. Tey, M. Seyedmahmoudian, M.Y. Idna Idris, S. Mekhilef, B. Horan, A. Stojcevski, SVR-based model to forecast PV power generation under different weather conditions, Energies 10 (7) (2017) 876.

[16] A. Mellit, A.M. Pavan, V. Lughi, Deep learning neural networks for short-term photovoltaic power forecasting, Renew. Energy 172 (2021) 276–288.

[17] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (8) (1997) 1735–1780, http://dx.doi.org/10.1162/neco.1997.9.8.1735.

[18] V. De, T.T. Teo, W.L. Woo, T. Logenthiran, Photovoltaic power forecasting using LSTM on limited dataset, in: 2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), IEEE, 2018, pp. 710–715.
[19] K. Wang, X. Qi, H. Liu, A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network, Appl. Energy 251 (2019) 113315.
[20] D. Lee, K. Kim, Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information, Energies 12 (2) (2019) 215.
[21] M. Castangia, A. Aliberti, L. Bottaccioli, E. Macii, E. Patti, A compound of feature selection techniques to improve solar radiation forecasting, Expert Syst. Appl. 178 (2021) 114979.
[22] S. Bozinovski, A. Fulgosi, The influence of pattern similarity and transfer learning upon the training of a base perceptron b2 (original in Croatian), in: Proceedings of Symposium Informatica, Bled, Slovenia, 1976.
[23] S. Zhou, L. Zhou, M. Mao, X. Xi, Transfer learning for photovoltaic power forecasting with long short-term memory neural network, in: 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), IEEE, 2020, pp. 125–132.
[24] X. Luo, D. Zhang, X. Zhu, Combining transfer learning and constrained long short-term memory for power generation forecasting of newly-constructed photovoltaic plants, Renew. Energy 185 (2022) 1062–1077, http://dx.doi.org/10.1016/j.renene.2021.12.104.
[25] Z. Yu, Z. Yang, F. Frank, H. Rao, W. Wie Song, Power Generation Prediction of Residential Photovoltaic Equipment Based on Online Transfer Learning Model-A Case Study of a Residential Solar Power System, Association for Computing Machinery, 2021, pp. 58–65.
[26] L. Bottaccioli, E. Patti, E. Macii, A. Acquaviva, GIS-based software infrastructure to model PV generation in fine-grained spatio-temporal domain, IEEE Syst. J. 12 (3) (2017) 2832–2841.
[27] T.D.S.C. LLC, Dark sky API, 2019, URL https://darksky.net/dev, Accessed on: 2019-06-01.
[28] W. Holmgren, C. Hansen, M. Mikofski, Pvlib python: a python package for modeling solar energy systems, J. Open Source Softw. 3 (2018) 884, http://dx.doi.org/10.21105/joss.00884.
[29] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (Mar) (2003) 1157–1182.
[30] A. Aliberti, Machine Learning Techniques to Forecast Non-Linear Trends in Smart Environments (Ph.D. thesis), Politecnico di Torino, 2020.
[31] I.H. Witten, E. Frank, M.A. Hall, C. Pal, M. DATA, Practical machine learning tools and techniques, in: Data Mining, Vol. 2, 2005, p. 4.
[32] C.A. Gueymard, A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects, Renew. Sustain. Energy Rev. 39 (2014) 1024–1034.
[33] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: International Joint Conference on Artificial Intelligence, 1995, URL https://api.semanticscholar.org/CorpusID:2702042.
[34] S. Shrivastava, Cross validation in time serie, 2023, URL https://medium.com/@soumyachess1496/cross-validation-in-time-series-566ae4981ce4, Accessed on: 2023-09-25.
[35] S. contributors, Using k-fold cross-validation for time-series model selection, 2023, URL https://stats.stackexchange.com/questions/14099/using-k-fold-cross-validation-for-time-series-model-selection, Accessed on: 2023-09-25.
[36] D. Soekhoe, P. Putten, A. Plaat, On the impact of data set size in transfer learning using deep neural networks, in: International Symposium on Intelligent Data Analysis, 2016, pp. 50–60, http://dx.doi.org/10.1007/978-3-319-46349-0_5.
[37] R. Ribani, M. Marengoni, A survey of transfer learning for convolutional neural networks, in: 2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), IEEE, 2019, pp. 47–57, http://dx.doi.org/10.1109/SIBGRAPI-T.2019.00010.