

Research Paper

Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods

Mutaz AlShafeey*, Csaba Csáki

Department of Information Systems, Institute of Informatics, Corvinus University of Budapest, Budapest, Fővám tér 13-15, H-1093, Hungary

ARTICLE INFO

Article history:

Received 17 March 2021

Received in revised form 12 September 2021

Accepted 31 October 2021

Available online xxxx

Keywords:

Solar energy

Photovoltaic technology

Prediction model

Multiple regression

Artificial neural network

Prediction accuracy

ABSTRACT

As Photovoltaic (PV) energy is impacted by various weather variables such as solar radiation and temperature, one of the key challenges facing solar energy forecasting is choosing the right inputs to achieve the most accurate prediction. Weather datasets, past power data sets, or both sets can be utilized to build different forecasting models. However, operators of grid-connected PV farms do not always have full sets of data available to them especially over an extended period of time as required by key techniques such as multiple regression (MR) or artificial neural network (ANN). Therefore, the research reported here considered these two main approaches of building prediction models and compared their performance when utilizing structural, time-series, and hybrid methods for data input. Three years of PV power generation data (of an actual farm) as well as historical weather data (of the same location) with several key variables were collected and utilized to build and test six prediction models. Models were built and designed to forecast the PV power for a 24-hour ahead horizon with 15 min resolutions. Results of comparative performance analysis show that different models have different prediction accuracy depending on the input method used to build the model: ANN models perform better than the MR regardless of the input method used. The hybrid input method results in better prediction accuracy for both MR and ANN techniques, while using the time-series method results in the least accurate forecasting models. Furthermore, sensitivity analysis shows that poor data quality does impact forecasting accuracy negatively especially for the structural approach.

© 2021 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Irrespective of all the advantages of utilizing photovoltaic (PV) technology for energy production, there are some hindrances limiting growth and wider utilization. One of the crucial drawbacks of solar solutions is low energy converting efficiency (Huang et al., 2013). Other drawbacks come from the nature of solar radiation which, unfortunately, highly fluctuates over time leading to generation uncertainty. Fluctuation and uncertainty in energy production lead to uncertainty in economic benefits. Calculating economic indicators such as energy pricing, rate of return, and payback period is challenging under generation uncertainty. Additionally, uncertainty can affect grid stability in case of grid-connected PV farms (Alshafeey and Csáki, 2019).

There are many solutions that have been used to overcome the above problems, but most of them are either costly (such as batteries Koochi-Fayegh and Rosen, 2020) or maybe impractical in several situations (e.g. hybrid diesel generators Yamegueu et al., 2011; Cavalcante et al., 2021). One promising solution is

to enhance solar energy forecasting (Singh, 2013; Devaraj et al., 2021). If potential solar energy can be accurately predicted with lower uncertainty, solar systems can be better designed and optimized helping grid operators in managing power supply and demand (Pazikadin et al., 2020). Accurate forecasts would improve grid stability as well (Rodríguez et al., 2018). Improving the forecasting models is one of the most important aspects of solar energy production and is considered to be one of the 'hottest' topics in the solar energy research field. Solar energy prediction models are software solutions that can be used to forecast future values of solar power generation. The forecasting horizon is the time between the present and the effective time of the predictions, while forecasting resolution is the frequency of the predictions (Antonanzas et al., 2016). Like any system that predicts the future, the forecasted value of energy would have a degree of uncertainty and errors. A good prediction model can forecast future values with minimum errors and uncertainties (Cammarano et al., 2012).

A potential problem with such forecasting models concerns the ability of grid-connected PV farm operators to utilize their advantages. Operators may not be familiar with the different techniques and hiring relevant expertise might be costly, but more importantly, they often face a situation when they do not have

* Correspondence to: Budapest, Fővám tér 13-15, H-1093, Hungary.
E-mail addresses: Mutaz.AlShafeey@uni-corvinus.hu (M. AlShafeey),
Csaki.Csaba@uni-corvinus.hu (C. Csáki).

all necessary data available to train such models successfully. Therefore, the main aim of the research reported here was to clarify the effects of utilizing different input data configurations on the forecasting accuracy in order to help PV grid-connected farms to achieve better forecasting accuracy within the available data that is normally collected from PV farms. Different PV farms have different data depending on the sensors installed or their budget (to purchase data from 3rd party). The study presented here aimed to evaluate different methods for building PV solar power forecasting models each utilizing different combinations of real PV solar power, geographical, and meteorological variables as input. As a result, three methods of utilizing input data – structural, time-series and hybrid – were used to build statistical (multiple regression - MR) and machine learning (ANN) prediction models. The datasets used for building and testing the models were collected from a 546 kWp grid-connected solar farm. The performance of the six models utilizing different configurations of the available input data were then evaluated. Hence grid operators can consider the best model based on the data available to them. This manuscript expands on prior research conducted and published in the field, yet the novelty of this work is derived from three aspects. First, prior researches either focus on one modelling technique utilizing different input data methods or focus on different modelling techniques utilizing one input data method – however, this work focuses on both. Second, the design and test of the models are the results of a systematic approach followed by a comprehensive performance analysis over several measures. And finally, the dataset is an outcome of the operation of a grid-connected PV farm, thus the testing results are not simulated, instead, they are outcomes based on large amount of real-life data under various weather conditions over several years.

The paper is organized as follows. After the introduction, PV solar energy forecasting methods are reviewed. This is followed by an overview of the methodology applied where the source data and model approaches used are described. The main section deals with the evaluation of various methods, models, and their settings – resulting in some potential improvement options. The paper closes with conclusions and future research.

2. PV solar energy forecasting methods

Solar power forecasting is a sophisticated process, many factors affect its accuracy. Other than the choice or combination of modelling methods, selecting forecast horizon and forecast model inputs (prediction variable patterns) are key decisions to be made (Ahmed et al., 2020). In addition, there are several measures established to show performance and compare accuracy of solutions (these would include various forms of error calculations among others and are discussed in detail in Section 3.3.3).

The forecasting horizon is the time between the present and the effective time of the predictions (Antonanzas et al., 2016). Although thus far there is no international classification criterion (Nespoli et al., 2019; Sobri et al., 2018), most studies categorize forecasting horizon into three classes: short-term (up to 48–72 h ahead), medium-term (from a few days to one week ahead), and long-term forecasting (from a few weeks to a year or more ahead) (Mellit et al., 2020; Raza and Khosravi, 2015; de Marcos et al., 2019). The second factor that affects the prediction model accuracy is the set of input variables. Choosing the right variables is one of the challenges in designing PV power forecasting models (Ahmed et al., 2020). Selecting input variables imprudently increases cost, computational complexity, and forecasting errors (Raza et al., 2016).

One of the simplest techniques to forecast solar energy is using the average values of historical solar energy and weather

records (Abunima et al., 2019). However, the average method is not suitable because averages do not represent the full range of values, which will be reflected by having some considerable errors in the forecasted values. Furthermore, those errors and the inherent uncertainty of solar radiation will be aggravated, leading to additional uncertainty (Linguet et al., 2016).

Two competing key techniques for solar power forecasting are statistical calculations and machine learning (Das et al., 2018). Machine Learning (ML) is the study of computer algorithms that are able to improve automatically through ‘experience’. ML models can solve complex problems by establishing complex relations between inputs and outputs and their performance usually depends on the quality and quantity of data available for training (Harrington, 2012). They have been used for many purposes including classification, pattern recognition, spam detection, and they are also used in data mining and forecasting problems (Voyant et al., 2017). Machine learning techniques are divided into supervised, unsupervised, and meta-learning algorithms (Lantz, 2019). Each of these algorithms has one or more learning tasks, i.e. supervised learning algorithms can be used for numeric prediction and classification problems, while unsupervised learning algorithms can be used for pattern detection and clustering (Lantz, 2019). One of the supervised learning algorithms is Artificial Neural Networks (ANN). The ANN technique can solve complex nonlinear, nonanalytical, and nonstationary stochastic problems without complex programming required (Inman et al., 2013).

The above abilities and advantages have influenced many researchers to use ANN in solar forecasting – including irradiation and energy production prediction (Wang et al., 2019a). As a result, a rising number of research reports and ANN-based forecasting applications have been published in the past 20 years (Garud et al., 2021). For example, ANN is used to forecast day-ahead solar energy with 1-hour forecasting resolutions utilizing historical weather data as well as time-series power data (Chen et al., 2011). ANN has also been applied as a base for short-term solar power prediction models (Almonacid et al., 2014) that showed a good performance for 1-hour power forecasting. Three different short-term prediction models using ANN were built in Oudjana et al. (2013). The first model utilized temperature data to forecast power and showed huge errors; the second model utilized solar irradiance data which resulted in better forecasting accuracy; while the third model showed the best accuracy and it utilized both temperature and solar irradiance to forecast power. ANN PV power forecasting model based on a self-organizing feature map (SOFM) was proposed in Yousif et al. (2017), where the suggested model uses solar irradiance and ambient temperature to forecast PV power. The results show that using ANN based on SOFM improves prediction accuracy. Real-time solar irradiance was used to make two-hour-ahead solar irradiance levels forecasting in Vanderstar et al. (2018): the proposed method uses ANN to forecast the irradiance and genetic algorithm to optimize array size and position in order to obtain the most accurate prediction. The suggested method shows adequate forecasting capabilities, yet, it has some limitations as this method only works for non-zero solar Global Horizontal Irradiance (GHI) values. In Notton et al. (2019) ANN models were proposed to forecast different solar irradiance components for 1 to 6-hour horizons. The results show that ANN is a very promising method to forecast solar radiation. Several ANN forecasting models were proposed to predict hourly solar irradiance in six different locations in Nigeria (Bamisile et al., 2020). The results show that all of the proposed ANN models performed well and can be used for PV performance calculation. Multiple weather variables such as temperature, precipitation, wind speed, and solar irradiation were used to build a multi-channel convolutional neural network (CNN) prediction model in

Table 1
Forecasting methods, horizons, resolutions, and variables.

Reference	Forecast horizon	Forecast resolution	Methods	Variables
Oudjana et al. (2012)	7 days	24 h	Linear regression, MR, neural network	Global irradiance, temperature
Al-Messabi et al. (2012)	10 and 60 min.	10 and 60 min	Dynamic ANN	Actual and past values of power
Ogliari et al. (2013)	24 hours-	1 h	ANN hybrid approach	Weather variables
De Giorgi et al. (2014)	1–24 h	1–24 h	Statistical methods based on MR analysis; ANN	PV power, module temperature, ambient temperature, solar irradiance
Chu et al. (2015)	5–15 min.	5 min.	Many methods including cloud tracking, k-NN, ANN	Power past values and sky images
Leva et al. (2017)	24 hours-	1 h	ANN	Power and solar radiation past values, Numerical Weather Prediction variables
Pitalúa-Díaz et al. (2019)	30 days	5 min	MR, Gradient Descent Optimization (GDO) and Adaptive Neuro-Fuzzy Inference System (ANFIS)	Solar radiation, ambient temperature, wind speed, daylight hour and PV power

Heo et al. (2021). The suggested model extracts meteorological as well as geographical features of PV sites from raster image datasets. The results show high forecasting capabilities, however, to avoid any biased prediction, sufficient data should be included. ANN was the most applied technique for solar power forecasting over the last ten years especially for short-term prediction as 48% of related articles published between 2009–2019 were using ANN (Mellit et al., 2020).

Unlike ML methods which formulate solar energy prediction problems as a black box, statistical methods reveal the mathematical relationship between the input variables and the output (Wang et al., 2020a). Such statistical methods include Autoregressive Moving Average (ARMA), Auto-Regressive Integrated Moving Average (ARIMA), exponential smoothing, and regression (Wang et al., 2020b; Das et al., 2018). Multiple Linear Regression (MLR) is also popular in PV solar power forecasting (e.g. De Giorgi et al., 2014; Oudjana et al., 2012; Pitalúa-Díaz et al., 2019). Regression methods establish a relationship between the explanatory (meteorological and geographical) variables and dependent variables (the forecasted PV power) (Das et al., 2018). Table 1 provides a brief chronological overview of the main PV forecasting methods applied for different horizons using various resolutions and input variables.

To forecast solar power, prediction models may utilize different parameters as inputs. Choosing among potential variables depends on the availability of related data for the required location and time period (as some locations do not have full datasets covering all the required parameters and for an extended period of time). Depending on the parameters used – i.e. the so-called explanatory variables utilized – the prediction models can be built using three different approaches (Aggarwal and Saini, 2014; Bacher et al., 2009; Jafarzadeh et al., 2012; Khatib et al., 2012):

- structural methods that only utilize the geographical and meteorological parameters as inputs;
- time-series methods that only utilize the historical data of solar power as inputs; and
- hybrid methods that utilize solar power historical data as well as other variables like geographical and meteorological parameters as inputs.

It should be noted, that there are two basic approaches to time-series forecasting: direct forecasting and multi-step rolling forecasting. While in the direct approach only actual historical data is utilized (i.e. always being one-time horizon behind), in the rolling approach the predictions of the previous values are used like they were actual values when predicting the next value (being one resolution step behind). Although there are some claims that multi-step rolling forecasting is slightly better than the direct option for certain tasks (see for example Lan et al.,

2019 for frequency-based solar irradiation forecasting), for solar power output this method has been found to be problematic. This is because the error generated in each step is propagated to the subsequent steps (Sahoo et al., 2020). Thus, it is found to be less accurate due to the accumulation of the error along the prediction horizon (Galicia et al., 2019). Consequently, this paper focuses on direct time-series forecasting.

While ANN and MLR are heavily studied methods, most of the literature focus either on testing one method of modelling utilizing different variables (see for example Inman et al., 2013; Oudjana et al., 2013) or on testing different methods utilizing the same variables (such as Ogliari et al., 2013; Chu et al., 2015; De Giorgi et al., 2014), the goal of this study was to compare and analyse the performance of two different modelling techniques – namely ANN and multiple regression – for PV power forecasting each using all three above-mentioned methods of utilizing input data. For instance, ANN solar radiation forecasting model was proposed and benchmarked with MLR and ARIMA models using satellite-derived land-surface temperature (LST) as an input predictor for 21 stations in Queensland (Deo and Şahin, 2017). Even though the study uses satellite-derived data to enhance solar radiation forecasting accuracy, yet the authors suggested future studies to test for shorter time-scale, as well as utilizing satellite data merged with reanalysis and ground-based products. Another study evaluated the PV power forecasting performance of different ANN models and some simple statistical models like ARMA, ARIMA, and Seasonal Auto-Regressive Integrated Moving Average (SARIMA) (Sharadga et al., 2020), yet, only time series of the power output was utilized. In study (Qu et al., 2021) time-series data of PV power and some related weather variables namely radiation, temperature, and humidity were utilized in different combinations to build a hybrid neural network forecasting model (composed of attention-based, long-term, and short-term temporal parts, i.e. ALSM). The proposed model shows better forecasting accuracy than ARMA and ARIMA models. However, only one input data method was utilized (hybrid time-series input data). Hence, compared with existing research on similar topics this paper provides a comprehensive comparison of PV solar power forecasting options and systematically evaluates the effectiveness and applicability of deterministic predictors over two different methods. Moreover, the main goal of this research was to assist operators of PV grid-connected farms to achieve better forecasting accuracy depending on the data available to them in an industrial situation. Solar farm operators do not necessarily have full sets of data that includes (all) geographical and meteorological parameters as well as past power, especially over an extended period of time. Therefore, it is essential for them to be able to judge the expected accuracy they may achieve depending on the data available and to make smart decision which method (or methods) to use.

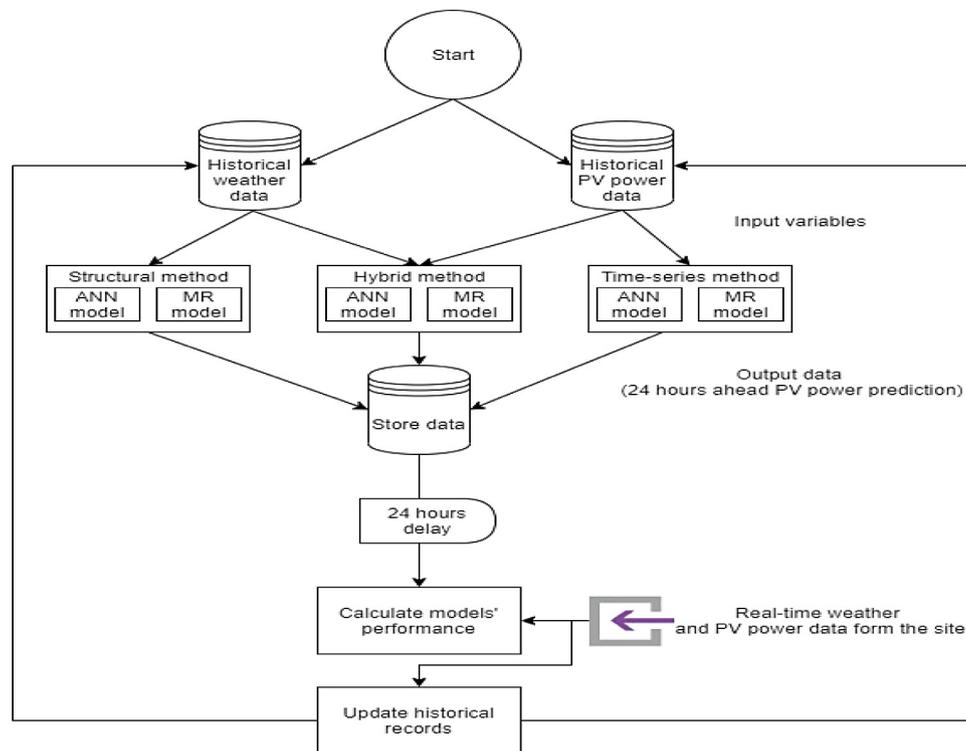


Fig. 1. General overview of the methodology flowchart.

Source: Authors.

3. Design of the experiment and methods

3.1. Objective and research methodology

Key challenges facing solar energy forecasting models include the task of choosing the right method and the need to select appropriate inputs to achieve the most accurate prediction. Consequently, the research reported here aimed at investigating two of the main techniques for building prediction models to accurately forecast PV output production: multiple regression (MR) and artificial neural network (ANN). To that end, structural, time-series, and hybrid data input methods were used to build different forecasting models and experiment with different input (predictor) settings.

Fig. 1 depicts a general overview of the steps forming the development process. Building the forecasting models starts by feeding the historical weather and PV power data to the models. Structural models are fed with historical weather data, time-series models are fed with historical PV power data, while hybrid models are fed with both weather and PV power historical data. Each model is forecasting the PV output power for the selected horizon with a given resolution set (for details see next Subsection). The forecasted PV output power values are then stored. When the real values become available as a fact data from the PV farm, this data is used to calculate the performance of each model (i.e. in comparison to the stored prediction) as well as to update the historical data records (which means this real-time data is later applied to update the model).

As part of the research, a large amount of historical data was collected to build the prediction models. The data collected covered the period April 13, 2017 to April 18, 2020 (3 years). The data used to train the models for prediction is measured data. Past data is used as an input to forecast the next day, albeit differently depending on input method. For example the measurement of April 13 had been used to predict expected output for April 14

in case of the structural model, and for April 15th in cases of the time-series and hybrid models. In other words, for the structural model weather data from exactly 24 h earlier is used to forecast for a given point in time – e.g. any timeslot of April 14 may be predicted using data from April 13. However, for the time series data prediction (and, therefore, for the hybrid model as well) a full past day data of generated power is needed as input for the model to forecast the next day – e.g. a prediction done on April 14 to predict the same timeslot on April 15 (one-day-ahead) uses data covering a full 24 h going back (thus including data from April 13), consequently, no prediction is possible for April 14 if data is not available from April 12. This was repeated until April 18, 2020. Thus, all models were continuously trained and tested over the data covering a 3-year period. For all ANN models trained here, the data was split into three segments: 70% training 15% validation, and 15% test set. Since this implies tens of thousands of values of each variable, it would be hard to visualize the forecasted versus the real power values for the whole period. Therefore, the last day of testing (18th of April 2020) was used to visualize and compare the performance of the different models. This day appeared to be a good test day as it had a few dips during the day due to weather changes during the day (as opposed to an average stochastic PV power curve).

3.2. Data and data collection

To achieve the aims of this study, large amount of data is required to build and test the proposed prediction models. Thus, various sources were used to collect data. Two international firms were collaborating with the authors to provide the necessary historical and real-time data. A Hungarian operator of several large farms provided the authors with full access to the power generation data of a grid-connected PV farm using SolarEdge farm management interface. While geographical and meteorological data for the solar farm location were provided by a specialized firm offering solar data services, Solcast Technologies. Table 2 shows the

Table 2
Technical site information.

Location	Hungary
Date of the establishment	April 13th, 2017
Peak power kWp	546
Data obtained	Energy and power generated (every 15 min)
PV cell model	ND-RJ260
Number of modules	2100
Number of inverters	18

PV solar farm site specifications including location, peak power, established date, data obtained, and other technical information.

Cell temperature was not provided by the data source, hence it was calculated using Eq. (1) (Mattei et al., 2006; Trinuruk et al., 2009):

$$T_c = T_a + (T_s - 20) * \frac{\text{solar irradiance}}{800} \quad (1)$$

where T_c , T_a , and T_s are the cell, ambient, and the Standard Test Conditions (STC) temperature in Celsius. The T_s for the PV models used for this study is 25°.

Historical weather data records include air temperature, cloud opacity, dewpoint temperature, Diffuse Horizontal Irradiance (DHI), Direct Normal Irradiance (DNI), Direct Beam Horizontal Irradiance (EBH), Global Horizontal Irradiance (GHI), Global Tilted Irradiance (GTI) fixed-tilt, GTI Tracking, precipitable water, relative humidity, snow depth, and wind speed. These variables are described in Table 3 (all collected with 15 min resolutions between April 13, 2017, and April 18, 2020).

3.3. Major factors affecting solar power forecast

This section presents the forecast horizons, forecast model inputs, and performance estimation measures used in this study.

3.3.1. Forecast horizon

Forecasting horizon is one of the major factors that affect forecasting accuracy (Das et al., 2018). The relationship between the forecasting horizon and accuracy is reverse: forecasting accuracy decreases significantly for longer horizons (Ahmed et al., 2020). As this research was designed to help grid-connected PV farms to achieve better forecasts within the available data, the models are intended to forecast 24 h ahead as the day-ahead PV energy forecasting is of the utmost importance in decision-making processes (Cococcioni et al., 2011). Moreover, certain grid operators in the European Union (EU) are required to report a 24-hours-ahead with 15 min resolution forecast from each grid-connected PV farm (Orasch, 2009; Zsiborács et al., 2019).

3.3.2. Forecast model inputs

PV power output is highly correlated with weather variables (Wang et al., 2019b). Yet, not all weather variables have the

Table 4
The correlation between PV output power and meteorological variables.

Input variables	Correlation with PV output power
Air temperature	0.42
Cloud opacity	−0.26
Dewpoint temperature	0.18
DHI	0.68
DNI	0.84
EBH	0.87
GHI	0.95
GTI fixed tilt	0.96
GTI tracking	0.90
Precipitable water	0.09
Relative humidity	−0.53
Snow depth	−0.09
Wind speed 10 m	0.09
Cell temperature	0.52

same significance for PV power forecasting. A correlation analysis was done to determine the significance of each collected variable shown in Table 3 before it was used in the modelling: see Table 4 for results.

As can be seen in Table 4, some meteorological factors have higher significance than others. Solar irradiance components have the highest significant factors, especially GTI fixed-tilt. Generally, it can be concluded that all variables collected here can be used in the modelling, yet variables that have low correlations with the output power could be excluded: in this study the threshold is set to 0.1, therefore wind speed, snow depth, and precipitable water are excluded.

3.3.3. Performance calculations

Subsequently, a comparative performance analysis between the resulting models shall be executed in order to compare prediction results and identify the most accurate model depending on specified evaluation criteria. To evaluate the performance of the forecasting model, one or more evaluation methods are needed. Accuracy of the forecasting models can be evaluated using Mean Absolute Error (MAE), Mean Squared Error (MSE), Coefficient of Determination (COD), Error (ϵ), and Root Mean Square Error (RMSE). Table 5 summarizes those methods (Ahmed et al., 2020; Elsheikh et al., 2019), where n is the number of observations, y_t and p_t are the observed (real) and the forecasted output power values at time t , respectively, and \hat{y} is the average of the observed values.

MAE is a quantity that is used in order to measure the closeness of the predicted values to the measured values. MSE measures the average of the squares of the errors – and thus embodies not only how widely the estimates are spread from the real sample but also how far off the average estimated value is from the true value. COD has been used to show how close prediction model results are to the actual measured data line as a fitted regression line (also known as R-squared (R^2) score and is generally used for testing hypotheses). Error is the actual (not

Table 3
Variables used in the study.

Name	Type	Unit	Source
Air temperature	Historical weather	Celsius	Solcast
Cell temperature	Calculated	Celsius	Calculations
Wind speed	Historical weather	m/s	Solcast
Cloud opacity	Historical weather	Percentage (%)	Solcast
Dewpoint temperature	Historical weather	Celsius	Solcast
DHI, EBH, DNI, GHI, and GTI	Historical weather	W/m ²	Solcast
Precipitable water	Historical weather	Centimetres	Solcast
Relative humidity	Historical weather	Percentage (%)	Solcast
Snow depth	Historical weather	Centimetres	Solcast
PV power generation	Historical power	W	SolarEdge
Forecasted PV power	Predicted variable	W	Prediction model output

absolute) value of the difference between the estimation and the corresponding actual value. RMSE is calculated to measure the prediction of a given approach thus shows the so-called scattering level produced by the model. For higher modelling accuracy, MAE, MSE, ε and RMSE indices should be closer to zero but the COD value should be closer to 1.

4. The PV power forecasting models used

Both forecasting techniques, multiple regression and ANN were applied to the dataset (as prepared in Section 3.2) using three different methods: structural, time-series, and hybrid – leading to six models in total. Regression models were developed using R packages, while ANN models were built in Matlab.

4.1. Multiple regression model and its three variants

A multiple linear regression model for PV power forecast can be denoted as follows:

$$y = \beta_0 + \beta_1 v_1 + \beta_2 v_2 + \beta_3 v_3 + \beta_4 v_4 + \dots + \beta_n v_n + \varepsilon \quad (2)$$

where v_1, v_2, \dots, v_n are the input variables (1 to n). The coefficient β_0 is the intercept, while values of β_1, \dots, β_n denote the slope coefficient of each input (explanatory) variable, and ε is the error (the amount by which the predicted value is different from the actual value). The regression model estimates the best values of β_1, \dots, β_n leading to least error ε .

Given all or part of the explanatory variables (Table 3), the multiple regression model can be used to forecast PV output power (y) for a given time t . To forecast future values of y within the forecast horizon (i.e. $y_{(t+1)}, y_{(t+2)}, \dots, y_{(t+h)}$, where h is the number of time periods between the present and the effective time of predictions), past values of v are required. Since the resolution is 15 min, to forecast one day ahead h is set to 96 time periods. Consequently, for a given time t , y is predicted using v values from time $(t - h)$. Notice, that as explained under methodology, forecasting should start on the second day of data collection to avoid negative times.

In structural method, the input variables fed to the multiple regression model are only the meteorological and geographical variables – with low correlation variables left out according to Section 3.3.2. Past values of selected weather variables were utilized according to Eq. (3) which denotes the Structural Multiple Regression model (SMR):

$$y(t) = \beta_0 + \beta_1 v_{1(t-h)} + \beta_2 v_{2(t-h)} + \beta_3 v_{3(t-h)} + \dots + \beta_n v_{n(t-h)} + \varepsilon \quad (3)$$

PV output power can be forecasted by knowing its past values (Cococcioni et al., 2011). Therefore, for the time-series method only PV output power values were used as input such that actual past PV output power values were used to predict future values ($y(t)$) of the output power according to Eq. (4) which denotes the Time-series Multiple Regression model (TMR):

$$y(t) = \hat{\beta}_0 + \hat{\beta}_1 y_{(t-h-0)} + \hat{\beta}_2 y_{(t-h-1)} + \dots + \hat{\beta}_h y_{(t-h-95)} + \varepsilon \quad (4)$$

where $\hat{\beta}_1, \dots, \hat{\beta}_h$ denote the slope coefficient of each input (explanatory) variable, exactly as explained earlier, the hat symbol used here just to indicate that each model has its unique beta values. $y_{(t-h-0)}, y_{(t-h-1)}, \dots, y_{(t-h-95)}$ are the past PV power values starting from 24 h before the forecasting takes place (i.e. 24 h before the 1st prediction) and goes until 48 h before the 1st prediction. In other words, $y_{(t-h-0)}$ is the past PV power value 24 h (96 time period) before forecasting $y(t)$, $y_{(t-h-1)}$ represents the past PV power value 97 time period before forecasting $y(t)$, and $y_{(t-h-95)}$ represents the past PV power value 191 time period before forecasting $y(t)$. So, for predicting $y(t)$, 96 past PV power values are utilized by the model. Note that to forecast $y(t)$, the

Table 5
Definitions of performance measures.

Measure	Equation
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} * \sum_{i=1}^n y_t - p_t $
Mean Squared Error (MSE)	$MSE = \frac{1}{n} * \sum_{i=1}^n (y_t - p_t)^2$
Coefficient of Determination (COD)	$R^2 = \frac{\sum_{i=1}^n (y_t - p_t)^2}{\sum_{i=1}^n (y_t - \hat{y})^2}$
Error	$\varepsilon = y_t - p_t$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} * \sum_{i=1}^n (y_t - p_t)^2}$

past PV power values 48 h prior to the forecast are required, thus the first day of forecasting PV output power by TMR model was April 15th, 2020. This ensures that there is no overlap between the prediction horizon period (of 24 h) and the data representing the actual power generated, which becomes available at the end of each day.

Finally, in the hybrid method past values of all the variables, including actual past values of the power variable as well as past values of the weather variables were fed to the model to predict future PV output power values as shown in Eq. (5) representing the Hybrid Multiple Regression model (HMR):

$$y(t) = \bar{\beta}_0 + \beta_1 v_{1(t-h)} + \beta_2 v_{2(t-h)} + \beta_3 v_{3(t-h)} + \dots + \beta_n v_{n(t-h)} + \hat{\beta}_1 y_{(t-h-0)} + \hat{\beta}_2 y_{(t-h-1)} + \dots + \hat{\beta}_h y_{(t-h-95)} + \varepsilon \quad (5)$$

where $\bar{\beta}_0$ is the intercept for the HMR model.

4.2. Artificial neural network (ANN) model and its three variants

ANN is a network of “neurons” that are arranged in a layered structure. Fig. 2 shows a simple diagram of ANN where input variables arrive from the bottom, while the forecasted variable(s) (output) appears at the top layer. An ANN also includes one or more hidden layers and hidden neurons. Such an ANN structure where the information flow is directed in one direction only (from the bottom to the top layer) is called Multi-Layer Feed-Forward Neural Network (MLFFNN). Note that in this work MLFFNN networks are used. In order to evaluate the performance of the training algorithm performance, the error (difference between the MLFFNN output and the real measured output) is determined using MSE equation in Table 5. To minimize MSE values between the observed real output and the forecasted output back-propagation algorithm is used where MSE values are utilized to update the weights and the biases of the network.

In MLFFNN, each layer receives its inputs from the previous layer (except for the input layer), so the outputs of a certain layer are the inputs to the next one. Inputs to each neuron in a given (hidden or output) layer are combined using a weighted linear combination. Then a nonlinear activation (transfer) function φ modifies the results before it is ready to be output. This network structure has many advantages for this forecasting context as this structure works well with big data and provides quick predictions after training. Moreover, it can be applied to solve complex non-linear problems and same accuracy ratio can be achieved even with even smaller data (Khishe et al., 2018; Akkaya and Çolakoglu, 2019). Fig. 3 shows the flow of information in one artificial neuron, where w_1, w_2, \dots, w_n are the weights corresponding to input data v_1, v_2, \dots, v_n respectively (Kim, 2017).

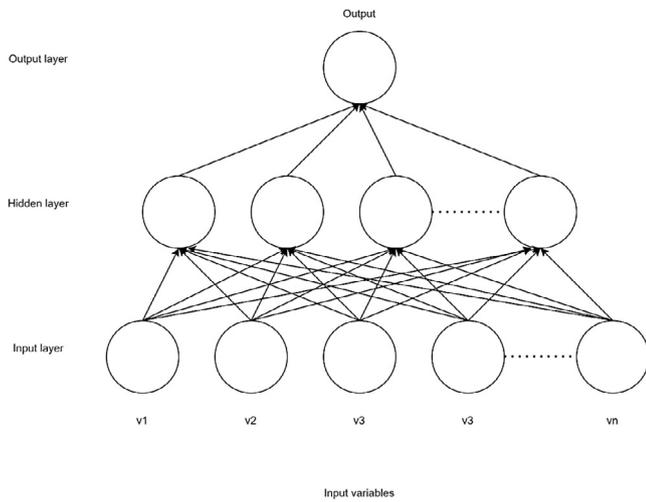


Fig. 2. A neural network with n inputs and one output. Source: Authors.

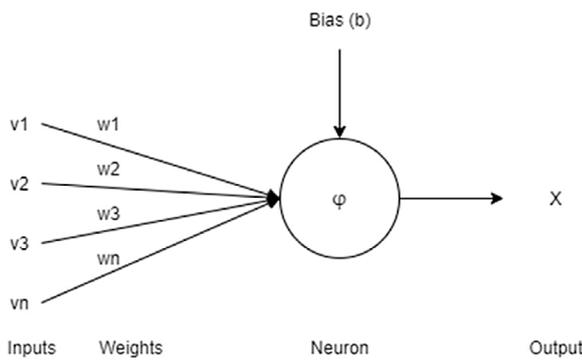


Fig. 3. Flow of information in an artificial neuron. Source: Authors.

Before reaching the neuron, the input signal of each neuron from the previous layer is multiplied by its dedicated weight. Once the weighted signals are collected, they are added to create the weighted sum (w_s) as denoted by Eq. (6), where b is the bias for the neuron:

$$w_s = w_1 \times v_1 + w_2 \times v_2 + \dots + w_n \times v_n + b \tag{6}$$

The weighted sum equation can be written with matrices as in (7) (Kim, 2017):

$$w_s = wv + b \tag{7}$$

where w and v are defined as:

$$w = [w_1 \ w_2 \ w_3 \ \dots \ w_n] \text{ and } v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ \vdots \\ v_n \end{bmatrix} \tag{8}$$

Then finally, the neuron enters the weighted sum into the activation function and yields its output as shown in Eq. (9):

$$X = \varphi(w_s) \tag{9}$$

One of the most used activation functions is the sigmoid function (Ghritlahre and Prasad, 2018) as given by (10):

$$\varphi(w_s) = \frac{1}{1 + e^{-w_s}} \tag{10}$$

Note that the sigmoid function is replaced by the Tansig transfer function whenever negative values are found in the input or output layers (Ghritlahre and Prasad, 2018) as follows:

$$\varphi(w_s) = \frac{1 - e^{-2w_s}}{1 + e^{-2w_s}} \tag{11}$$

The above equations represent one neuron but each individual neuron has its own specific set of weights and bias on the inputs. The weights of neurons are initially set to random values. Training data is fed to the bottom (input) layer and it passes through the succeeding layers, getting multiplied and added together as described in the equations, until it finally arrives, drastically transformed, at the output layer. Information is stored in form of weights. Those weights have to be changed to train the ANN with new information. In this work back-propagation algorithm is applied to minimize the MSE between the real observed output and forecasted output from the MLFFNN, the weights are adjusted in proportion to the input value(s) and the output error (MSE) as can be seen in (12) (Talaat et al., 2020).

$$\text{Min}(MSE) = \min\left(\frac{1}{n} * \sum_{i=1}^n (y_t - p_t)^2\right) \tag{12}$$

The change in weights and biased are calculated as in (13) and (14) respectively (Talaat et al., 2020; Leema et al., 2016):

$$\Delta w_n = \gamma(y_n - p_n) \tag{13}$$

$$\Delta b_n = \gamma(y_n - p_n) \tag{14}$$

where Δw_n is the change of weight for the n th neuron, Δb_n is the change of the bias for the n th neuron, and γ is the learning rate. Subsequently, the adjusted weights ($w_{adjusted}$) and biases ($b_{adjusted}$) are denoted as in (15) and (16) respectively.

$$w_{adjusted} = w + \Delta w \tag{15}$$

$$b_{adjusted} = b + \Delta b \tag{16}$$

One unit of this process (when training data is passed forward through the neural network and then the weights of each neuron are adjusted based on the error) is called an epoch. During training, the weights are repeatedly adjusted each epoch. This loop will continue until either a specific number of epochs is reached or when the value of MSE reaches the lowest possible limit (typically when MSE does not change for several epochs).

This research utilizes a fully connected MLFFNN as described above with one hidden layer. For each of the three methods used, the ANN was fed with the same set of input variables as were the corresponding MR models. Notice, this implies a differing number of input neurons for each method used. The number of input neurons, therefore, are 11, 96, and 107 for the structural, time series, and hybrid input methods respectively. The number of hidden neurons is an important parameter for ANN. With few hidden neurons the ANN might not be able to generate a function that indicates the underlying problem while having more hidden neurons than required may result in over-fitting of the training set and reducing the ability of generalization the out-of-sample data (Setyawati, 2005). Therefore the number of hidden neurons was set to be 33% (one-third) of the number of inputs. Table 6 shows the settings of the ANN parameters. Although training time had not been limited, the actual running time for the set number of epochs to be trained was ranging from seconds to a couple of hours, while forecasting times were, of course, very short (fraction of a second).

Table 6
ANN parameters.

Parameter	Description	Value for each method		
		Structural	Time series	Hybrid
Number of inputs	Number of input data variables	11	96	107
Number of outputs	Number of output forecasted variables	1	1	1
Number of hidden neurons	Number of hidden neurons	4	32	35
Maximum epochs	Maximum number of training iterations before training is stopped	1000	1000	1000
Maximum training time	Maximum time in seconds before training is stopped	Unlimited	Unlimited	Unlimited
Performance goal	The minimum target value of MSE	0	0	0

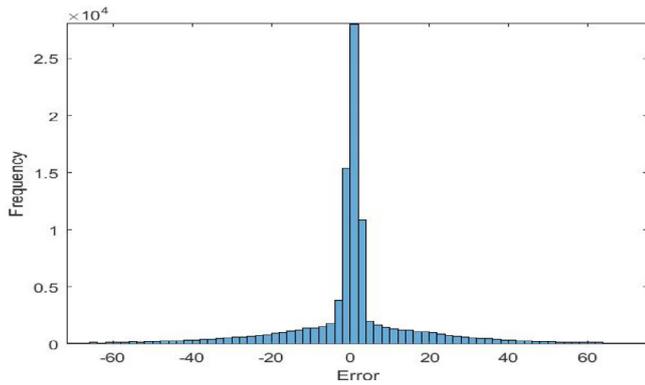


Fig. 4. Frequency distribution of the error in SMR.
Source: Authors.

5. Results of the experiments and discussion

After developing the suggested models, a series of experiments were constructed to measure the performance of each model variant and to compare their prediction abilities using the evaluation measures described in Table 5. The overall (average) performance of each variant was computed at the end of training and testing (i.e. over 3 years). Then, as an additional demonstration that enables some representative visualization, the performance for forecasting the output power for 18th of April 2020 was also computed and compared to the overall performance (see Table 7).

5.1. Multiple regression models

Initially, the multiple regression model was developed utilizing only meteorological and geographical variables, then another MR model was built to utilize only time-series data of PV solar power, and finally, a third model was developed utilizing both PV power historical data as well as geographical and meteorological parameters as inputs (for details see Section 4.1).

5.1.1. Structural Multiple Regression model (SMR) performance

Over the training and testing period, this model shows a good performance with a 0.94 COD, 14.84 MAE, 1054.74 MSE, and, 32.47 RMSE. Fig. 4 shows the frequency distribution of the error. It can be seen that the most frequent errors recorded are small errors ranging between -20 and +20 kW. However, some fairly large errors can also be observed.

The SMR model was able to forecast the energy for the 18th of April with a 0.92 COD, 22.44 MAE, 1815.52 MSE, and, 42.60 RMSE performance measures. Which is a bit less than the overall performance. Fig. 5 shows the forecasted vs. observed power for the mentioned day.

Note that the model utilizes real values, that is they have 100% accuracy. This explains the very good prediction performance. In case meteorological parameters can only be provided with some degree of uncertainty, the SMR might have less accurate

Table 7
Performance measures comparison.

Model		Performance measures			
		COD (R^2)	MAE	MSE	RMSE
MR	SMR	0.94	14.84	1054.74	32.47
	TMR	0.68	45.83	5584.5	74.72
	HMR	0.95	16.05	835.68	28.90
	Average	0.86	25.57	2491.64	45.36
ANN	SANN	0.95	13.13	943.53	30.26
	TANN	0.75	36.38	4329.87	36.38
	HANN	0.96	13.52	914.10	30.23
	Average	0.89	21.01	2062.5	32.29

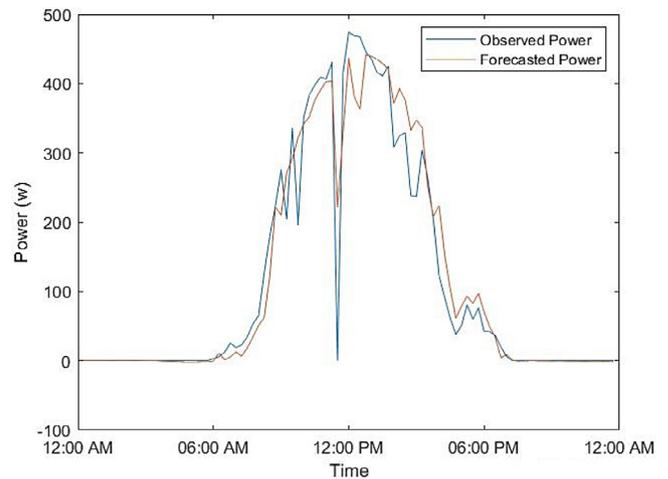


Fig. 5. Forecasted vs. observed power for the SMR.
Source: Authors.

performance. Fig. 6 shows a sensitivity analysis for the SMR model where the effect of uncertainty in the input variables can be observed on the forecasted power.

The sensitivity analysis shows the normalized percent changes in the forecasted PV output power with the normalized percent of input variables uncertainty. The point (1,1) on the graph represent 100% accurate inputs, therefore, there is no change in forecasted power. An uncertainty between 0 and ± 40% in any of the input variables (i.e. 0.6 and 1.4 in the x-axis) leads to huge changes in the model’s output, thus affecting the performance.

5.1.2. Time-series Multiple regression model (TMR) performance

Over the training and testing period, this model shows poor performance, much worse than the SMR with a 0.68 COD, 45.83 MAE, 5584.5 MSE, and, 74.72 RMSE. Fig. 7 shows the frequency distribution of the error. Some huge errors were recorded, additionally, the errors are not distributed around zero (most frequent errors do not equal zero)

Fig. 8 shows the forecasted vs. observed power for the 18th of April 2020. It can be noticed from error measures and Fig. 8 that the TMR performs worse than the SMR. The TMR could not

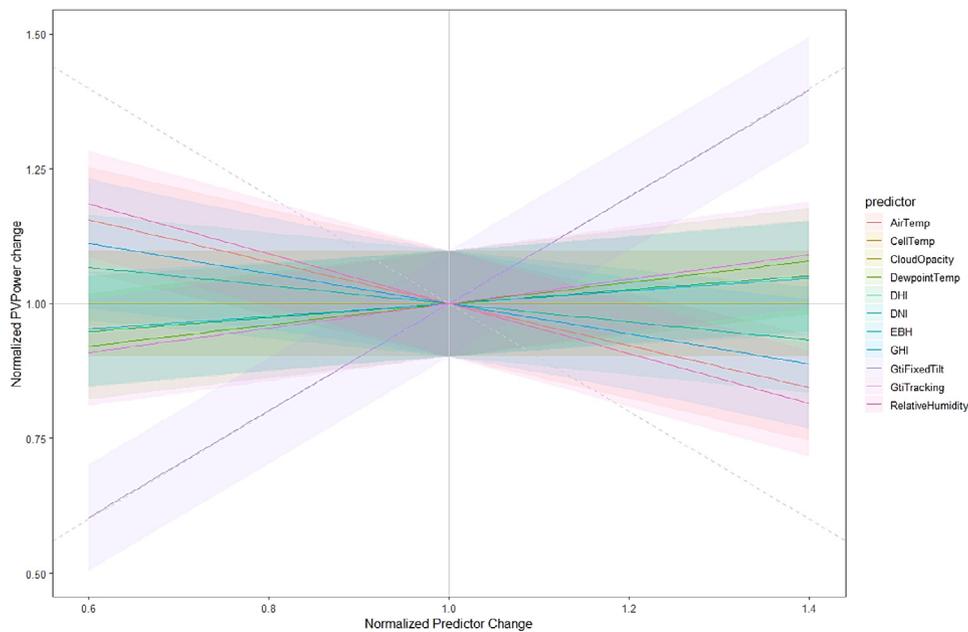


Fig. 6. Sensitivity analysis for the SMR model. Source: Authors.

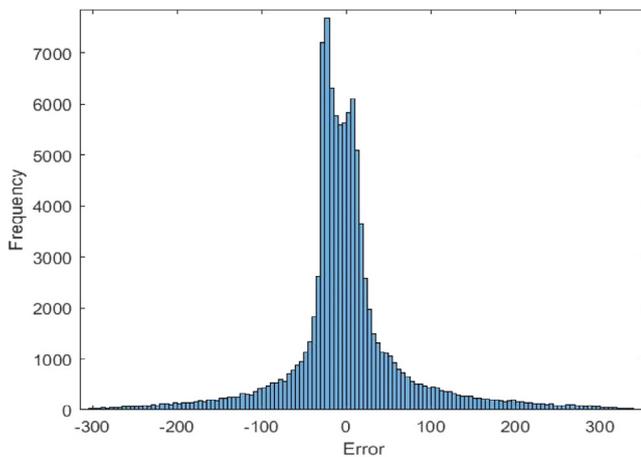


Fig. 7. Frequency distribution of the error in TMR. Source: Authors.

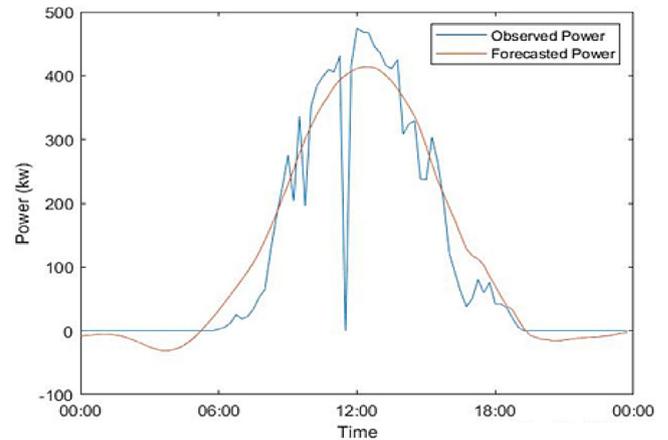


Fig. 8. Forecasted vs. observed power for the TMR. Source: Authors.

predict the sudden drop in the output PV power just before noon, while this drop was better predicted by the SMR.

The SMR model was able to forecast the energy for the 18th of April with a 0.88 COD, 32.97 MAE, 3091.27 MSE, and, 55.59 RMSE performance measures. Which is above the overall performance of this model. This can be explained as the 18th of April does not have huge weather variations, thus the TMR performs better than other days.

5.1.3. Hybrid Multiple regression model (HMR) performance

Over the training and testing period, this HMR model shows the best overall performance compared to TMR and SMR, with a 0.95 COD, 16.05 MAE, 835.68 MSE, and, 28.90 RMSE. Fig. 9 shows the frequency distribution of the error. Even though the HMR shows better performance, yet it shows more errors between -50 and 50 kW.

The HMR model was able to forecast the energy for the 18th of April with 0.93 COD, 22.26 MAE, 1796.84 MSE, and, 42.38

RMSE performance measures. Which is a bit less than the overall performance. Fig. 10 shows the forecasted vs. observed power for that day.

5.2. Artificial neural network models

The ANN models were built the same way as were the MR models in Section 5.1. Initially, the ANN model was developed utilizing only selected meteorological and geographical variables, then another ANN model was developed to utilize only time-series data of PV power, and finally, a third ANN model was developed to utilize PV power historical data as well as geographical and meteorological parameters as inputs.

5.2.1. Structural Artificial Neural Network model (SANN) performance

Over the training and testing period, this model shows a good performance with a 0.95 COD, 13.13 MAE, 943.53 MSE, and, 30.26 RMSE. The SANN reached the best performance (least MSE) after

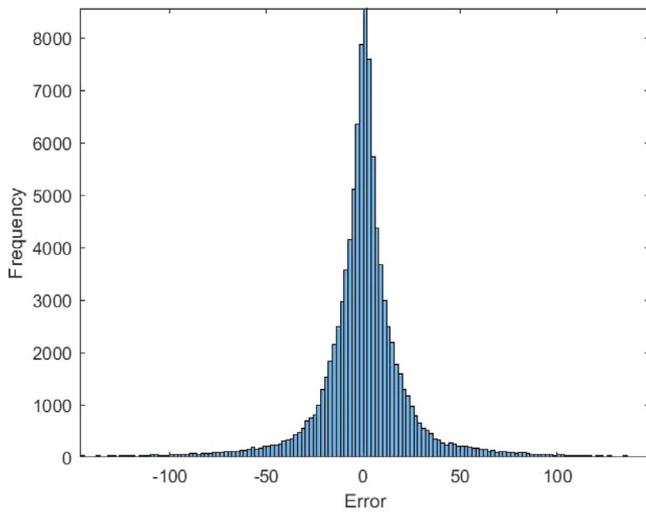


Fig. 9. Frequency distribution of the error in HMR.
Source: Authors.

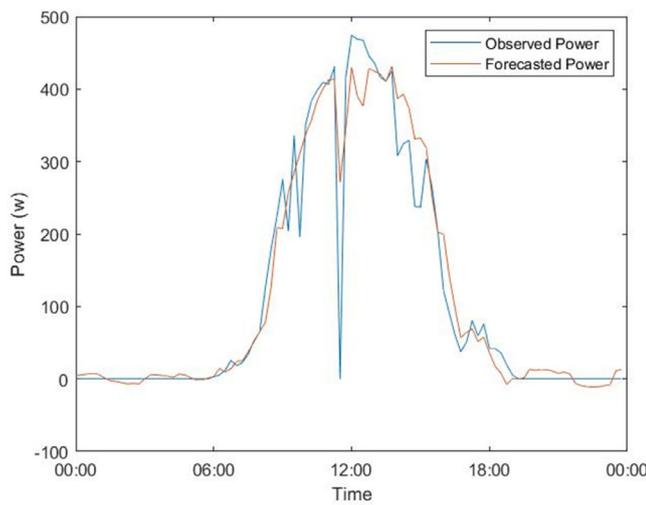


Fig. 10. Forecasted vs. observed power for the HMR.
Source: Authors.

140 epochs as shown in Fig. 11. The error distribution is in Fig. 12 which shows the total frequency of errors as well as the error frequency in the training, validation, and test sets.

The TMR model was able to forecast the energy for the 18th of April with 0.93 COD, 20.96 MAE, 1752.54 MSE, and, 41.86 RMSE performance measures. Which is a bit less than the overall performance. Fig. 13 shows the forecasted vs. observed power for that day.

5.2.2. Time-series Artificial Neural Network model (TANN) performance

Over the training and testing period, this model shows a fair performance, slightly better than the TMR with a 0.75 COD, 36.38 MAE, 4329.87 MSE, and, 36.38 RMSE. The TANN reached the best performance (least MSE) after 241 epochs (Fig. 14). The error distribution can be found in Fig. 15.

The TANN model was able to forecast the energy for the 18th of April with 0.87 COD, 32.57 MAE, 3135.32 MSE, and, 55.99 RMSE performance measures. Which is better than the overall performance but no dip is predicted, for the same reason mentioned in 5.1.2. Fig. 16 shows the forecasted vs. observed power for the mentioned day.

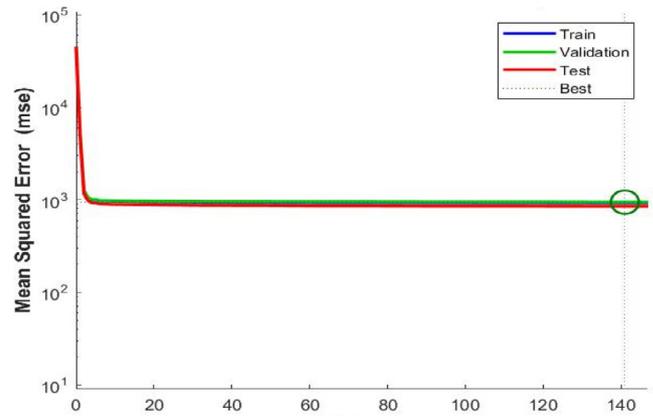


Fig. 11. SANN performance.
Source: Authors.

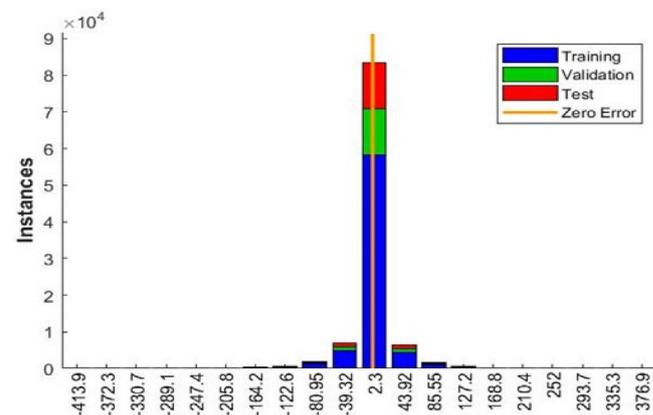


Fig. 12. Frequency distribution of the error in SANN.
Source: Authors.

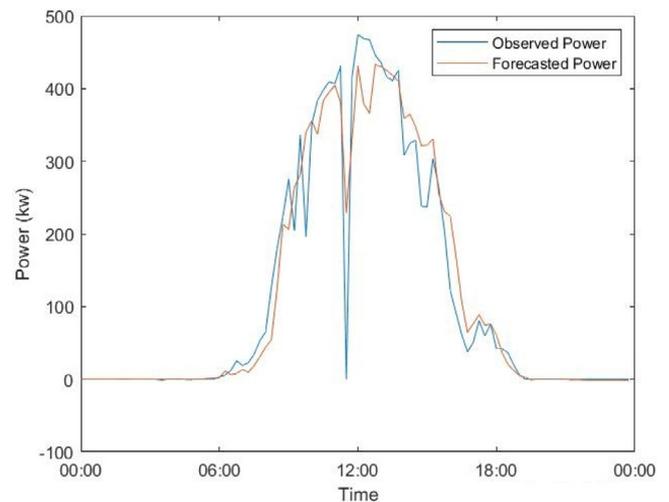


Fig. 13. Forecasted vs. observed power for the SANN.
Source: Authors.

5.2.3. Hybrid Artificial Neural Network model (HANN) performance

Over the training and testing period, the model shows a good performance, way better than the TANN with a 0.96 COD, 13.52 MAE, 914.10 MSE, and, 30.23 RMSE. The HANN reached the best performance (least MSE) after 29 epochs as shown in Fig. 17. The error distribution can be found in Fig. 18.

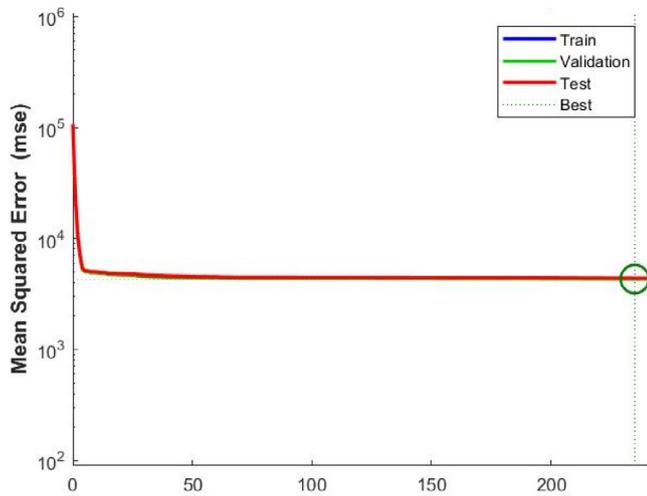


Fig. 14. TANN performance.
Source: Authors.

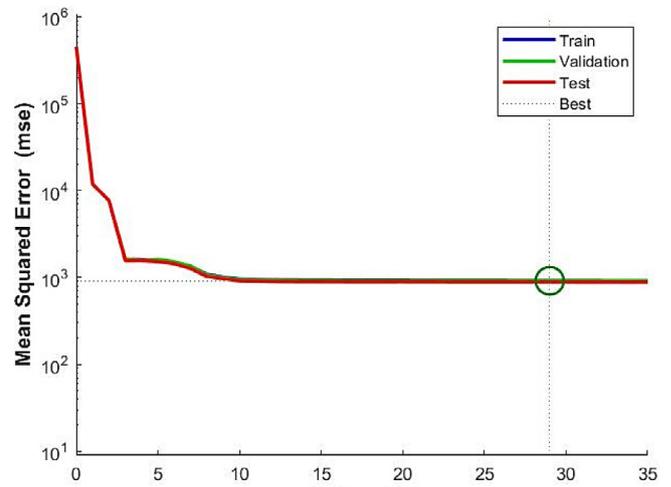


Fig. 17. HANN performance.
Source: Authors.

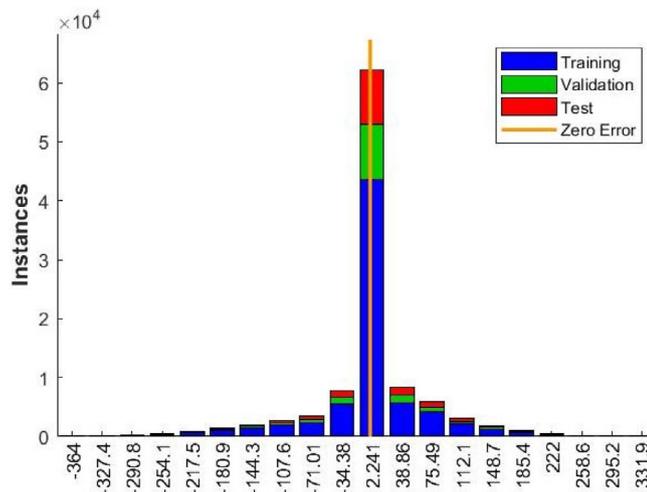


Fig. 15. Frequency distribution of the error in TANN.
Source: Authors.

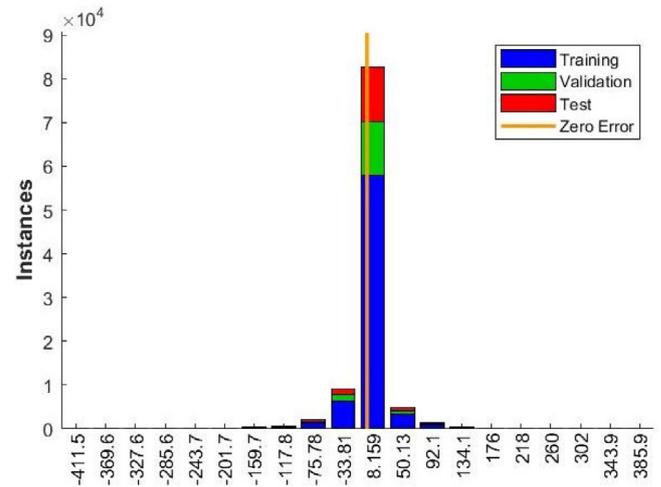


Fig. 18. Frequency distribution of the error in HANN.
Source: Authors.

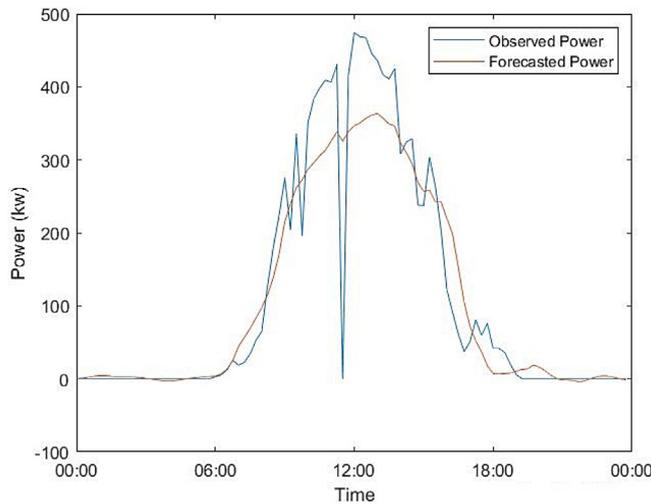


Fig. 16. Forecasted vs. observed power for the TANN.
Source: Authors.

The HANN model was able to forecast the energy for the 18th of April with 0.94 COD, 19.0 MAE, 1626.35 MSE, and, 40.32 RMSE performance measures. Which is almost the same as the expected performance. Fig. 19 shows the forecasted vs. observed power for the mentioned day with the daily dip predicted.

5.3. Performance comparison

In this section, a performance comparison for all the models that were designed and tested in the previous subsections is provided. Table 7 summarizes the overall performance data of the different models.

As Table 7 shows, the difference between MR and ANN is very clear in the time-series data, where TANN performance is highly superior compared to TMR. Moreover, even though SMR and SANN show comparable performances, yet the SMR is sensitive to the uncertainty in the input variables as discussed in the sensitivity analysis of Section 5.1.1. HANN has the highest COD, and the lowest MSE, and RMSE, thus the HANN has the best overall performance in all the measures used except MAE where SANN has the lowest value. It can be noticed that ANN is generally overperformed MR models – as can also be observed from the

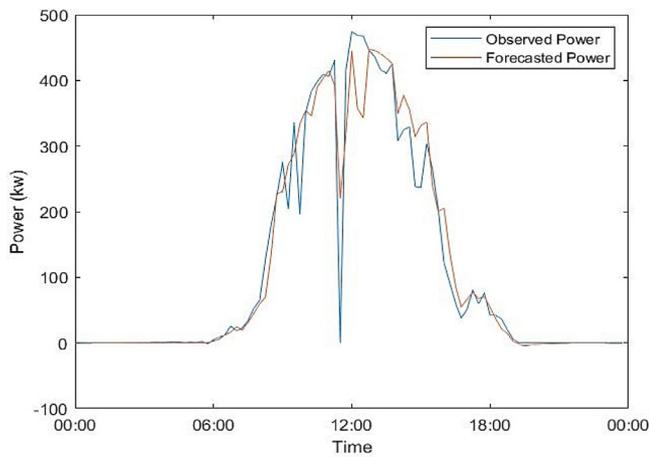


Fig. 19. Forecasted vs. observed power for the HANN.
Source: Authors.

diagrams in Fig. 20 which show average overall performance for the MR and ANN.

Fig. 21 shows average performance measures comparison between structural, time-series, and hybrid methods. Time-series models have the worst performances. Although hybrid and structural models have close performance values, hybrid models perform vaguely better in the MSE and RMSE measures, while structural models overperform the hybrid ones in the MAE measure.

When comparing the performance of the models for forecasting PV output power on April 18th, 2020 similar conclusions can be made, as expected: ANN models were slightly overperforming MR ones (see Table 8).

Here again, time-series models show unfavourable performance compared to the structural or hybrid models. The hybrid models show the highest COD and least MAE, MSE, and RMSE.

6. Conclusions

This paper demonstrates two different techniques of PV energy prediction modelling, namely ANN and MR. Depending on

the input variables utilized, forecasting models were built using three different approaches: structural, time-series, and hybrid. The six models were built to predict the PV solar power for a 546 kWp grid-connected solar farm located in Hungary. This research is targeted to help PV farms improving their power prediction, therefore, the horizon and the resolution of the forecasts were set based on the forecasting regulations affecting certain grid operators in the European Union. Hence, all forecasting models were built and designed to forecast PV output power for a 24-hour ahead horizon with 15 min resolution. So, a historical data set including 3-years of geographical and meteorological variables was collected for the site of this specific PV farm along with actual PV power values. This data was used to build, train, and test the models.

The results indicate that ANN forecasting models have higher COD and lower MAE, MSE, RMSE values compared to the MR, regardless of the method used for building the forecasting models. It was also found that using the hybrid method to build prediction models results in better prediction accuracy for both MR and ANN while using the time-series method results in the least accurate forecasting models.

After analysing the results of this work using real farm data, it was confirmed that ANN technique performs better than the MR. This is true regardless of the input method used to build the models. It was also found that using the hybrid method of input data to build the forecasting models leads to better forecasting accuracy regardless of the technique used. The results of sensitivity analysis show that input variables and corresponding data quality have huge effects on the models' output when utilizing the structural technique. Consequently, in case of poor data quality or inaccurate weather data it is recommended to avoid using the structural method, especially when using the structural method to build MR forecasting models. To summarize, farm operators may have better results using ANN-based models with hybrid input approach.

Finally, some improvements might be done to expand this work. One possibility is to compare the performance of the tested forecasting models for different horizons and resolutions. In theory, the performance of the forecasting models is decreasing for longer forecasting horizons. It would be worthwhile to study the performance of the tested models for medium- and long-term forecasting horizons.

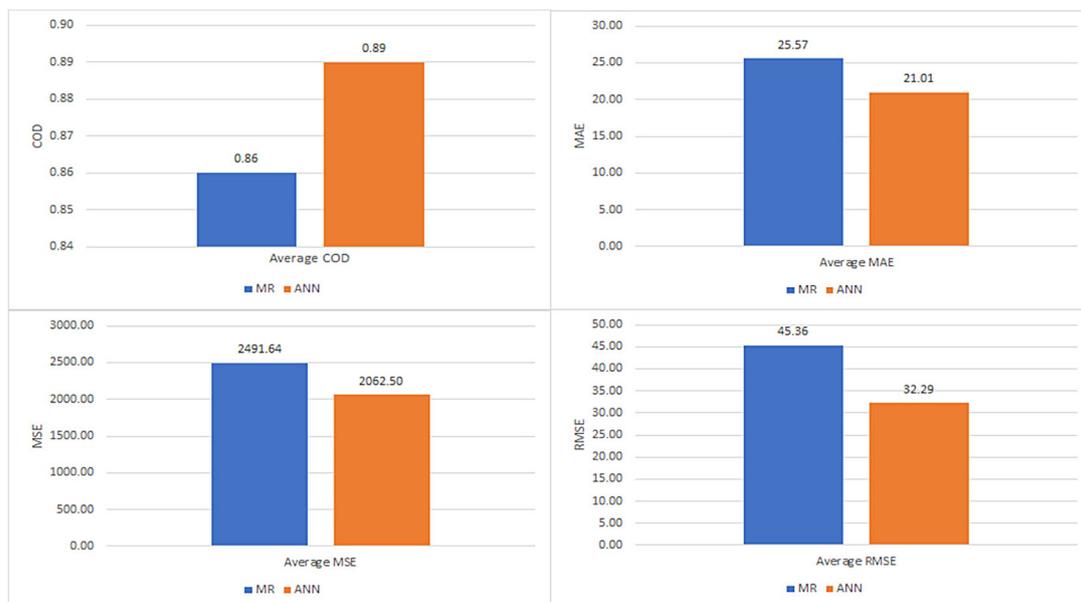


Fig. 20. Average performance measures comparison between MR and ANN.
Source: Authors.

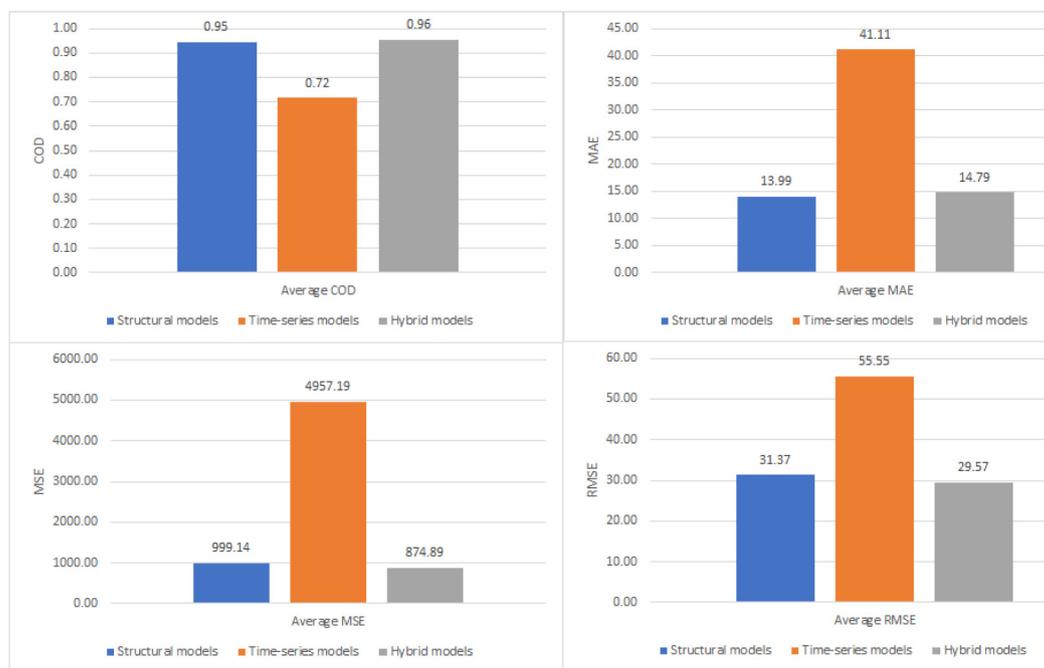


Fig. 21. Average performance measures comparison between structural, time-series, and hybrid models.
Source: Authors.

Table 8

Performance measures comparison for PV out power for the 18th of April 2020.

Model		Performance measures			
		COD (R^2)	MAE	MSE	RMSE
MR	SMR	0.92	22.44	1815.52	42.60
	TMR	0.88	32.97	3091.27	55.59
	HMR	0.93	22.26	1796.84	42.38
	Average	0.91	25.89	2234.54	46.86
ANN	SANN	0.93	20.96	1752.54	41.86
	TANN	0.87	32.57	3135.32	55.99
	HANN	0.94	19.0	1626.35	40.32
	Average	0.91	24.18	2171.40	46.06

CRedit authorship contribution statement

Mutaz AlShafeey: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – original draft. **Csaba Csáki:** Validation, Writing - original draft, Writing – review & editing, Methodology, Supervision.

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.

Acknowledgements

Our Artificial Application for Renewable Energy Production Prediction project is supported by solar irradiance and weather data from Solcast (<http://solcast.com>). Solar power plant operation data is provided by 3Comm Hungary (3comm.hu/). Project no. TKP2020-NKA-02 has been implemented with the support provided by the National Research, Development and Innovation Fund of Hungary, financed under the ‘Tématerületi Kiválóssági Program’ funding scheme.

References

- Abunima, H., Teh, J., Jabir, H.J., 2019. A new solar radiation model for a power system reliability study. *IEEE Access* 7, 64758–64766.
- Aggarwal, S., Saini, L., 2014. Solar energy prediction using linear and non-linear regularization models: A study on AMS (American Meteorological Society) 2013–14 Solar Energy Prediction Contest. *Energy* 78, 247–256.
- Ahmed, R., Sreeram, V., Mishra, Y., Arif, M., 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew. Sustain. Energy Rev.* 124, 109792.
- Akkaya, B., Çolakoglu, N., 2019. Comparison of Multi-class Classification Algorithms on Early Diagnosis of Heart Diseases. *y-BIS Conference 2019: Recent Advances in Data Science and Business Analytics*, Istanbul, Turkey, p. 162–72.
- Al-Messabi, N., Li, Y., El-Amin, I., Goh, C., 2012. Forecasting of photovoltaic power yield using dynamic neural networks. In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, pp. 1–5.
- Almonacid, F., Pérez-Higueras, P., Fernández, E.F., Hontoria, L., 2014. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. *Energy Convers. Manage.* 85, 389–398.
- Alshafeey, M., Csáki, C., 2019. A case study of grid-connected solar farm control using artificial intelligence genetic algorithm to accommodate peak demand. *J. Phys. Conf. Ser.* 012017, IOP Publishing.
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martínez-de Pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. *Sol. Energy* 136, 78–111.
- Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. *Sol. Energy* 83, 1772–1783.
- Bamisile, O., Oluwasanmi, A., Obiora, S., Osei-Mensah, E., Asoronye, G., Huang, Q., 2020. Application of deep learning for solar irradiance and solar photovoltaic multi-parameter forecast. *Energy Sour. Part A: Recovery, Util. Environ. Effects* 1–21.
- Cammarano, A., Petrioli, C., Spenza, D., 2012. Pro-energy: A novel energy prediction model for solar and wind energy-harvesting wireless sensor networks. In: *2012 IEEE 9th International Conference on Mobile Ad-Hoc and Sensor Systems (MASS 2012)*. IEEE, pp. 75–83.
- Cavalcante, R.L., Costa, T.O., Almeida, M.P., Williamson, S., Galhardo, M.A.B., Macêdo, W.N., 2021. Photovoltaic penetration in isolated thermoelectric power plants in Brazil: Power regulation and specific consumption aspects. *Int. J. Electr. Power Energy Syst.* 129, 106648.
- Chen, C., Duan, S., Cai, T., Liu, B., 2011. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. Energy* 85, 2856–2870.
- Chu, Y., Urquhart, B., Gohari, S.M., Pedro, H.T., Kleissl, J., Coimbra, C.F., 2015. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol. Energy* 112, 68–77.

- Cococcioni, M., D'Andrea, E., Lazerri, B., 2011. 24-hour-ahead forecasting of energy production in solar PV systems. In: 2011 11th International Conference on Intelligent Systems Design and Applications. IEEE, pp. 1276–1281.
- Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., et al., 2018. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* 81, 912–928.
- De Giorgi, M.G., Congedo, P.M., Malvoni, M., 2014. Photovoltaic power forecasting using statistical methods: impact of weather data. *IET Sci. Meas. Technol.* 8, 90–97.
- Deo, R.C., Şahin, M., 2017. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. *Renew. Sustain. Energy Rev.* 72, 828–848.
- Devaraj, J., Madurai Elavarasan, R., Shafiullah, G., Jamal, T., Khan, I., 2021. A holistic review on energy forecasting using big data and deep learning models. *Int. J. Energy Res.*
- Elsheikh, A.H., Sharshir, S.W., Abd Elaziz, M., Kabeel, A., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: A comprehensive review. *Sol. Energy* 180, 622–639.
- Galicia, A., Talavera-Llames, R., Troncoso, A., Koprinska, I., Martínez-Álvarez, F., 2019. Multi-step forecasting for big data time series based on ensemble learning. *Knowl.-Based Syst.* 163, 830–841.
- Garud, K.S., Jayaraj, S., Lee, M.Y., 2021. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. *Int. J. Energy Res.* 45, 6–35.
- Ghritlahre, H.K., Prasad, R.K., 2018. Application of ANN technique to predict the performance of solar collector systems—a review. *Renew. Sustain. Energy Rev.* 84, 75–88.
- Harrington, P., 2012. *Machine Learning in Action*. Manning Publications Co.
- Heo, J., Song, K., Han, S., Lee, D.-E., 2021. Multi-channel convolutional neural network for integration of meteorological and geographical features in solar power forecasting. *Appl. Energy* 295, 117083.
- Huang, X., Han, S., Huang, W., Liu, X., 2013. Enhancing solar cell efficiency: the search for luminescent materials as spectral converters. *Chem. Soc. Rev.* 42, 173–201.
- Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 39, 535–576.
- Jafarzadeh, S., Fadali, M.S., Evrenosoglu, C.Y., 2012. Solar power prediction using interval type-2 TSK modeling. *IEEE Trans. Sustain. Energy* 4, 333–339.
- Khatib, T., Mohamed, A., Sopian, K., 2012. A review of solar energy modeling techniques. *Renew. Sustain. Energy Rev.* 16, 2864–2869.
- Khishi, M., Mosavi, M., Moridi, A., 2018. Chaotic fractal walk trainer for sonar data set classification using multi-layer perceptron neural network and its hardware implementation. *Appl. Acoust.* 137, 121–139.
- Kim, P., 2017. *Matlab deep learning. With machine learning*. Neural Netw. Artif. Intell. 130, 21.
- Koohi-Fayegh, S., Rosen, M.A., 2020. A review of energy storage types, applications and recent developments. *J. Energy Storage* 27, 101047.
- Lan, H., Zhang, C., Hong, Y.-Y., He, Y., Wen, S., 2019. Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network. *Appl. Energy* 247, 389–402.
- Lantz, B., 2019. *Machine Learning with R: Expert Techniques for Predictive Modeling*. Packt Publishing Ltd.
- Leema, N., Nehemiah, H.K., Kannan, A., 2016. Neural network classifier optimization using differential evolution with global information and back propagation algorithm for clinical datasets. *Appl. Soft Comput.* 49, 834–844.
- Leva, S., Dolara, A., Grimaccia, F., Mussetta, M., Ogliari, E., 2017. Analysis and validation of 24 h ahead neural network forecasting of photovoltaic output power. *Math. Comput. Simulation* 131, 88–100.
- Linguet, L., Pousset, Y., Olivier, C., 2016. Identifying statistical properties of solar radiation models by using information criteria. *Sol. Energy* 132, 236–246.
- de Marcos, R.A., Bello, A., Reneses, J., 2019. Electricity price forecasting in the short term hybridising fundamental and econometric modelling. *Electr. Power Syst. Res.* 167, 240–251.
- Mattei, M., Notton, G., Cristofari, C., Muselli, M., Poggi, P., 2006. Calculation of the polycrystalline PV module temperature using a simple method of energy balance. *Renew. Energy* 31, 553–567.
- Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., Lughi, V., 2020. Advanced methods for photovoltaic output power forecasting: A review. *Appl. Sci.* 10, 487.
- Nespoli, A., Ogliari, E., Leva, S., Massi Pavan, A., Mellit, A., Lughi, V., et al., 2019. Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies* 12, 1621.
- Notton, G., Voyant, C., Fouilloy, A., Duchaud, J.L., Nivet, M.L., 2019. Some applications of ANN to solar radiation estimation and forecasting for energy applications. *Appl. Sci.* 9, 209.
- Ogliari, E., Grimaccia, F., Leva, S., Mussetta, M., 2013. Hybrid predictive models for accurate forecasting in PV systems. *Energies* 6, 1918–1929.
- Orasch, W., 2009. Regulatory framework for RES-e system integration in europe-description and analysis of different European practices. Appendix Deliver. 8.
- Oudjana, S.H., Hellal, A., Mahamed, I.H., 2012. Short term photovoltaic power generation forecasting using neural network. In: 2012 11th International Conference on Environment and Electrical Engineering. IEEE, pp. 706–711.
- Oudjana, S., Hellal, A., Mahammed, I.H., 2013. Power forecasting of photovoltaic generation. *Int. J. Electr. Comput. Eng.* 7, 627–631.
- Pazikadin, A.R., Rifai, D., Ali, K., Malik, M.Z., Abdalla, A.N., Faraj, M.A., 2020. Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): A review of five years research trend. *Sci. Total Environ.* 715, 136848.
- Pitalúa-Díaz, N., Arellano-Valmaña, F., Ruz-Hernandez, J.A., Matsumoto, Y., Alazki, H., Herrera-López, E.J., et al., 2019. An ANFIS-based modeling comparison study for photovoltaic power at different geographical places in Mexico. *Energies* 12, 2662.
- Qu, J., Qian, Z., Pei, Y., 2021. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* 232, 120996.
- Raza, M.Q., Khosravi, A., 2015. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* 50, 1352–1372.
- Raza, M.Q., Nadarajah, M., Ekanayake, C., 2016. On recent advances in PV output power forecast. *Sol. Energy* 136, 125–144.
- Rodríguez, F., Fleetwood, A., Galarza, A., Fontán, L., 2018. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renew. Energy* 126, 855–864.
- Sahoo, D., Sood, N., Rani, U., Abraham, G., Dutt, V., Dileep, A., 2020. Comparative analysis of multi-step time-series forecasting for network load dataset. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). IEEE, pp. 1–7.
- Setyawati, B.R., 2005. Multi-layer feed forward neural networks for foreign exchange time series forecasting. Department of Industrial and Management Systems Engineering. West Virginia University, Morgantown, West Virginia.
- Sharadga, H., Hajimirza, S., Balog, R.S., 2020. Time series forecasting of solar power generation for large-scale photovoltaic plants. *Renew. Energy* 150, 797–807.
- Singh, G.K., 2013. Solar power generation by PV (photovoltaic) technology: A review. *Energy* 53, 1–13.
- Sobri, S., Koohi-Kamali, S., Rahim, N.A., 2018. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manage.* 156, 459–497.
- Talaat, M., Farahat, M., Mansour, N., Hatata, A., 2020. Load forecasting based on grasshopper optimization and a multilayer feed-forward neural network using regressive approach. *Energy* 196, 117087.
- Trinuruk, P., Sorapipatana, C., Chenvidhya, D., 2009. Estimating operating cell temperature of BIPV modules in thailand. *Renew. Energy* 34, 2515–2523.
- Vanderstar, G., Musilek, P., Nassif, A., 2018. Solar forecasting using remote solar monitoring stations and artificial neural networks. In: 2018 IEEE Canadian Conference on Electrical & Computer Engineering (CCECE). IEEE, pp. 1–4.
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., et al., 2017. Machine learning methods for solar radiation forecasting: A review. *Renew. Energy* 105, 569–582.
- Wang, H., Lei, Z., Zhang, X., Zhou, B., Peng, J., 2019a. A review of deep learning for renewable energy forecasting. *Energy Convers. Manage.* 198, 111799.
- Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., et al., 2020a. Taxonomy research of artificial intelligence for deterministic solar power forecasting. *Energy Convers. Manage.* 214, 112909.
- Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., Shi, M., 2020b. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Convers. Manage.* 212, 112766.
- Wang, F., Zhang, Z., Liu, C., Yu, Y., Pang, S., Duić, N., et al., 2019b. Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers. Manage.* 181, 443–462.
- Yamegueu, D., Azoumah, Y., Py, X., Zongo, N., 2011. Experimental study of electricity generation by Solar PV/diesel hybrid systems without battery storage for off-grid areas. *Renew. Energy* 36, 1780–1787.
- Yousif, J.H., Kazem, H.A., Boland, J., 2017. Predictive models for photovoltaic electricity production in hot weather conditions. *Energies* 10, 971.
- Zsiborács, H., Hegedűsné Baranyai, N., Csányi, S., Vincze, A., Pintér, G., 2019. Economic analysis of grid-connected PV system regulations: A hungarian case study. *Electronics* 8, 149.