Energy 224 (2021) 120109

Contents lists available at ScienceDirect

Energy

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

Prediction of solar energy guided by pearson correlation using machine learning

Imane Jebli [a,*], Fatima-Zahra Belouadha [a], Mohammed Issam Kabbaj [a], Amine Tilioua [b]

[a] AMIPS Research Team, E3S Research Center, Computer Science Department, Mohammadia School of Engineers, Mohammed V University in Rabat, Avenue Ibn Sina B.P. 765, Agdal Rabat, 10090, Morocco

[b] Research Team in Thermal and Applied Thermodynamics (2.T.A.), Mechanics, Energy Efficiency and Renewable Energies Laboratory (L.M.3.E.R.), Department of Physics, Faculty of Sciences and Techniques Errachidia, Moulay Ismaïl University of Meknes, B.P. 509 Boutalamine, Errachidia, Morocco

ARTICLE INFO

Article history:
Received 4 May 2020
Received in revised form
11 February 2021
Accepted 13 February 2021
Available online 18 February 2021

Keywords:
Solar energy prediction
Machine and deep learning
Linear regression
Random forest
Support vector regression
Artificial neural networks

ABSTRACT

Solar energy forecasting represents a key element in increasing the competitiveness of solar power plants in the energy market and reducing the dependence on fossil fuels in economic and social development. This paper presents an approach for predicting solar energy, based on machine and deep learning techniques. The relevance of the studied models was evaluated for real-time and short-term solar energy forecasting to ensure optimized management and security requirements in this field while using an integral solution based on a single tool and an appropriate predictive model. The datasets we used in this study, represent data from 2016 to 2018 and are related to Errachidia which is a semi-desert climate province in Morocco. Pearson correlation coefficient was deployed to identify the most relevant meteorological inputs from which the models should learn. RF and ANN have provided high accuracies against LR and SVR, which have reported very significant errors. ANN has shown good performance for both real-time and short-term predictions. The key findings were compared with Pirapora in Brazil, which is a tropical climate region, to show the quality and reproducibility of the study.

1. Introduction

Nowadays, the world is turning towards the use of renewable energy to produce electricity and redefine the energy mix. Being able to introduce higher percentages of renewable electricity in the energy mix, is indeed crucial to create an ecological and durable electrical system [1]. In this context, solar energy has become one of the most promising renewable resources for power generation. However, the energy context is marked by the price variation, the demand variation, and the instability of renewable energy production. In the context of solar power energy, this instability depends on materials used for manufacturing Photovoltaic (PV) cells factors, but also, on other factors such as the non-stationary meteorological variables. Hence, it is important to use an appropriate renewable energy prediction model that can learn from historical meteorological data and help to optimize the operations and take appropriate decisions in the energy sector. This model would give advance information on power availability, which helps to ensure grid stability and to allow optimal unit commitment and economic dispatch [2].

In this context, it is worth noting that the accuracy of a predictive model depends on the relevance of the features (input variables) selected for performing the prediction, and also on the forecasting horizon with regard to the available data. In the solar energy area, it is hence important to be firstly, able to identify the meteorological variables that can help provide accurate projections. It is also important to use an integral solution (a single predictive model) that can help meet the different challenges and objectives associated with the different solar forecasting horizons.

PV forecasts range, indeed, from very short (real-time) to long-term horizons depending on the challenges and the targeted objectives. The real-time horizon is necessary for PV storage control and electricity marketing. The short-term is useful for decision making such as unit commitment and comprising of economic load dispatch, while the medium and the long terms are respectively essential for scheduling maintenance (actually based on the use of the IEC TS 62446 standard) and planning of PV plant [3,4]. Using multiple solutions (predictive models) and hence, different adapted datasets to perform the most optimal solar energy projections according to different forecasting horizons, still, a potential error

* Corresponding author.
E-mail address: jebli1imane@gmail.com (I. Jebli).

https://doi.org/10.1016/j.energy.2021.120109
0360-5442/

Nomenclature

Abbreviations Error Measures

ANN	Artificial Neural Network
MAE	Mean absolute error
LR	Linear Regression
MSE	Mean square error
RF	Random Forest
RMSE	Root mean square error
SVM	Support Vector Machine
ME	Max Error
SVR	Support Vector Regression
R2	R-Squared
MLP	Multilayer Perceptron
NRMSE	Normalized RMSE
PV	Photovoltaic

source that could be sanctioned by high-cost penalties. For example, in the case of grid energy management which is a crucial activity, the grid operators should frequently perform both real-time and short-term predictions. Thus, switching from a solution to another one when using specific and adapted data to each solution, increases the risk of human error and consequently, the risk of predicting error. Thus, the use of an integral solution is necessary to avoid human errors and also, to enable appropriate decision making based on a global visualization of different horizon predictions provided by the same predictive model.

In the literature, PV forecast models are classified into two categories: indirect and direct forecast models. As noted in Refs. [5,6], solar radiation on different time scales is forecasted using various methods, and then converted into power using the characteristics of panels in the case of the indirect forecast, while direct forecasts are made directly from the output power of the plant. Besides, solar energy prediction methods can be organized into four different classes: statistical approach [7,8], physical approach [9], artificial intelligence (AI) and, hybrid approach [10]. However, thanks to their high learning and regression capacities, AI and especially, Machine Learning (ML) techniques [11], have been used extensively in this field. According to Liu et al. [12], an RF PV prediction algorithm has been proposed using the Principal Component Analysis and K-means clustering algorithm. Wolff et al. [13] have evaluated several numerical weather prediction parameters to enhance PV power prediction capabilities. Lahouar et al. [14] have suggested a predictor based on RF, for predicting power ahead of daylight. The authors have integrated a prediction algorithm (Smart Persistence) using an advanced ML technique RF for PV power prediction implemented by Ref. [15]. Yao et al. [16] have used air quality index (AQI) as an extra input parameter to improve the performance of the SVM-2 models and estimate global solar radiation on a horizontal surface. Further, Fan et al. [17] have presented two ML methods, SVM and a novel simple tree-based ensemble method named extreme gradient boosting (XGBoost) to estimate solar radiation accurately with complete and incomplete temperature and precipitation data in humid subtropical climates. Their results show that the SVM and XGBoost models performed much better than the selected empirical models. Ayodele et al. [18] have implemented a hybrid algorithm based on k-means and Support Vector Regression (SVR) for predicting the daily global solar radiation. Vivek Palaniappan et al. [19] have used SVR for predicting solar radiation from historical meteorological data while using t-distributed Stochastic Neighbour Embedding to reduce the data dimensions. Hu et al. [20] have adopted a multi-layer BPNN to build a multi-model PV power prediction method based on seasonal characteristics in the Beijing area as input parameters. Honglu et al. [21] have introduced a method combining the advantages of ANNs (Artificial Neural Networks) and wavelet analysis used to carry out the multi-scale decomposition of the PV output. Cervone et al. [22] have presented a methodology based on ANN and an analog set in order to generate deterministic and probabilistic 72-h forecasts. Khan et al. [23] have presented the power prediction of a PV system 3 days in advance using a software engineering technique (Theoretical Solar Radiation) and Elman Neural Network. Alfadda et al. [24] have developed several ML models for 1-h advance solar irradiance forecasting, namely Multilayer Perceptron (MLP), Support Vector Regression (SVR), k-nearest neighbors (kNN), and decision tree regression. Yagli et al. [25] have tested a total of 68 ML models using global horizontal irradiance data obtained by satellite from 7 locations in 5 different climatic zones on the American continent for hourly solar forecasting. Paiva et al. [26] have evaluated two ML algorithms for predicting intraday solar irradiance, namely multi-genetic programming (MGGP) and the MLP for a short-term prediction horizon. Liu et al. [27] have adopted an ensemble technique using a recursive arithmetic average, to perform short-term stand-alone predictions of power output. Alizamir et al. [28] have studied and compared the capability of six algorithms, namely the gradient boosting tree (GBT), the MLP, two adaptive neuro-fuzzy inference systems (ANFIS) based on fuzzy c-means clustering (ANFIS-FCM) and subtractive clustering (ANFIS-SC), the multivariate adaptive regression spline (MARS), and the classification and regression tree (CART) for forecasting solar radiation. Deo et al. [29] have determined the accuracy of the ANN coupled with land-surface temperature derived from satellites as a monthly horizon predictor of solar radiation in the Queensland region. Feng et al. [30] have developed a back-propagation neural network using the Levenberg-Marquardt algorithm, to predict the monthly-mean daily global solar radiation. For this purpose, they used clouds, aerosols, and precipitable water-vapor data from Moderate Resolution Imaging Spectroradiometer along with conventional meteorological data. Furthermore, Ozoegwu [31] has predicted the monthly mean daily global solar energy based on time series with sufficient accuracy. They used a nonlinear autoregressive, a nonlinear autoregressive (exogenous), and a hybrid time series method implemented with ANNs.

All of these research contributions mentioned above are interesting. However, they don't propose an integral (unified and single) solution suitable for different forecasting horizons and thus, they aren't able to help meet the different challenges (PV system management, maintenance, and security) of the concerned field, while avoiding the complexity of using different tools. In addition, the studied contributions use different features as input variables including meteorological variables. However, they don't try to identify the most crucial parameters (features) that would really impact the prediction accuracy. Due to the non-linear behavior of the output power generation which varies from day to day depending on weather conditions such as temperature, wind speed, atmospheric aerosol levels, and humidity, the uncertainties of PV output power are increasing [32]. Hence, using all of the meteorological features or inappropriate ones could have a negative impact on the quality of the results and the accuracy of the models, as we will see in this paper.

In this context and given the importance of the management activities in the energy sector, the present work aims to propose a simple predictive model that is based on the most relevant meteorological parameters, to guarantee sufficient accuracies for both real-time and short-term forecasts of solar energy. It would be interesting for especially grid and market operators since it contributes to propose a unified and single tool that they could use for different real-time and daily needs ranging from real-time unit commitment to storage control forecasts, economic load dispatch, and electricity marketing. In other terms, using such a tool based on the most relevant input meteorological parameters is important since it would contribute to improve the PV system stability, reliability, and safety [33].

The main contributions and purposes of this study are as follows: (i) to show the efficiency of performing the cross-correlations between different meteorological variables using Pearson correlation coefficient, in order to identify the relevant meteorological data from which the models should learn to perform accurate predictions, (ii) to evaluate the efficiency of four machine and deep learning models for predicting solar energy using different metrics, and identify which model is able to provide a sufficient accuracy for both real-time and short-term solar energy predictions, (iii) to experiment our models with a semi-desert climate area, namely Errachidia region in Morocco, a country which hosts Noor, the largest solar power plant in the world, and which have initiated many renewable energy projects to obtain 42% of solar and wind electricity by 2020 and 52% by 2030, and (4i) to show the quality and reproducibility of our study by comparing the key findings with Brazil, especially Pirapora, a region with a tropical climate that is different from the semi-desert one, and which hosts the largest solar power plant in Latin America.

The remainder of this paper is organized into four sections. Section 2 provides an overview of the preliminaries, the main case study used in this work, and the data preparation. Subsequently, section 3 presents the methodology applied to elaborate the predictive models. Section 4 presents and discusses the obtained results. Finally, section 5 summarizes the conclusions and perspectives of this study.

2. Preliminaries and data preparation

2.1. Description of the grid-connected PV system

In general, the three main components of a grid-connected solar photovoltaic system are a photovoltaic generator that generates electricity directly from sunlight, power converters, and grid interface control [34]. The photovoltaic power relies on the power of the photovoltaic generator and the efficiency of the grid-connected inverter. For a specific PV system, the output power depends on the corresponding meteorological impact factors, including solar radiation, temperature, humidity, wind speed, and pressure. These parameters are potential for the estimation of this output power [35], as the production of photovoltaic energy is mainly based on the amount of solar irradiance [36]. Therefore, the received solar energy is measured by solar radiation defined as the amount of electromagnetic energy incident on a surface per unit time and per unit area [37].

In the area of solar energy forecasting, as a first step, the solar radiation model has to be set up. Then, according to the predicted value of historical local weather data, correlations between solar radiation and photovoltaic energy production can be established, in order to calculate the final predicted values [38]. Considering the effect of solar radiation and temperature, the output power of a PV panel is calculated from the following equation [39]:

$$P_{out,pv} = C_{R,PV}\frac{G_T}{G_{T,STC}}\left[1+\alpha_p(T_C - T_{C,STC})\right] \quad (1)$$

where, $C_{R,PV}$ is the rated capacity of the PV array (W), G_T and $G_{T,STC}$ are the solar radiation incident on the PV panel in the present time step and at standard test conditions (W/m^2), respectively. α_p is the temperature co-efficient of power (%/°C), while T_C and $T_{C,STC}$ are PV cell temperature in present time step and standard test conditions (°C), respectively.

2.2. Forecasting process

In the field of ML-based solar energy prediction, the adopted forecasting process usually consists of five steps: data acquisition, data pre-processing, feature extraction and identification, algorithm training, and algorithm testing to evaluate the model's performance. The architecture of the corresponding framework that we implemented in this study, is illustrated in Fig. 1. This framework processes temperature, solar radiation and energy, humidity, wind speed, wind direction and pressure.

Besides, we note that we use the mentioned framework to experiment it with the case of Errachidia. This province is located in a sunny part of Morocco, exactly in the south-east of this country (Fig. 2). It has a semi-desert climate, according to the Köppen-Geiger classification. The average temperature in this region is 20.1 °C, and the average rainfall is 156.9 mm.

This province is also characterized by these main features [41]:

- A significant thermal difference between very high summer temperatures (31.5 °C as average in July) and very low winter temperatures (5 °C as average in January).
- Weak and irregularly distributed rainfall in time and space. The majority of the territory receives less than 100 mm of rain per year.
- Winds recording speeds above 57.6 km/h in May, June July, and August.

In the next subsections, we give details about the main steps that we have performed according to the framework mentioned above, to forecast solar energy from data related to the Errachidia case study.

2.3. Meteorological datasets

As Errachidia city is located in the sunny region of Morocco, it benefits from an important solar energy potential during all the year. Bouaichi et al. [42] have provided the results of a first-time study on the long-term performance, degradation, and cost analysis of three different PV module technologies - monocrystalline silicon, polycrystalline silicon, and amorphous silicon installed in the semi-desert climate of Errachidia, Morocco. The datasets used in our article, represent meteorological data of Errachidia province: solar radiation, temperature, wind direction, wind speed, humidity, and pressure. The response parameter to be predicted is solar energy. This data includes measurements from three years (2016, 2017, and 2018), and is used to predict the solar energy level based on a set of measurable weather conditions. Knowing when conditions are most suitable for solar radiation and solar energy, is critical for deciding when and where to deploy solar energy recovery equipment.

2.4. Data pre-processing

We aim to model solar energy according to the available functionalities. It is therefore logical that solar energy is the label of the algorithm. The actual values used to train and test the supervised ML algorithm refer to registered solar energy measurements. Before applying ML techniques, the input data must be cleaned and normalized, as this historical data may contain different peaks and non-stationary components due to uncertainty and fluctuating

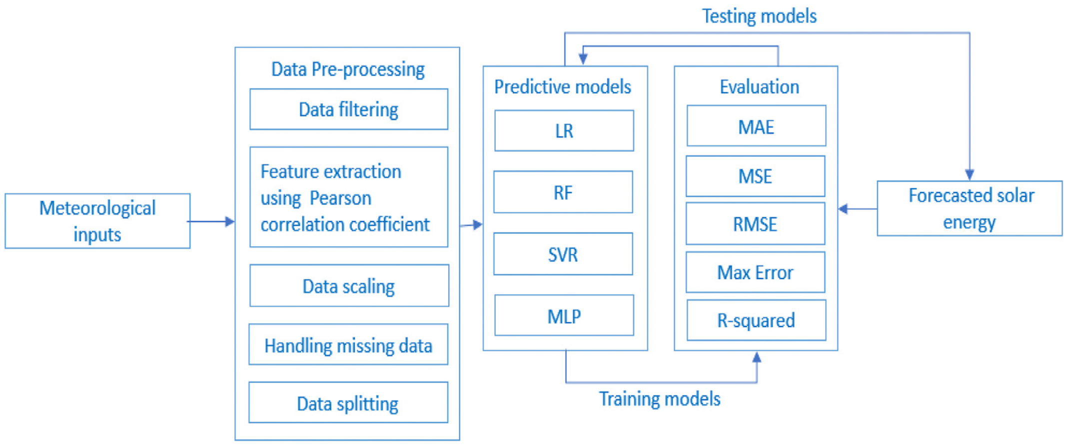

Fig. 1. Architecture of the proposed framework using ML algorithms.

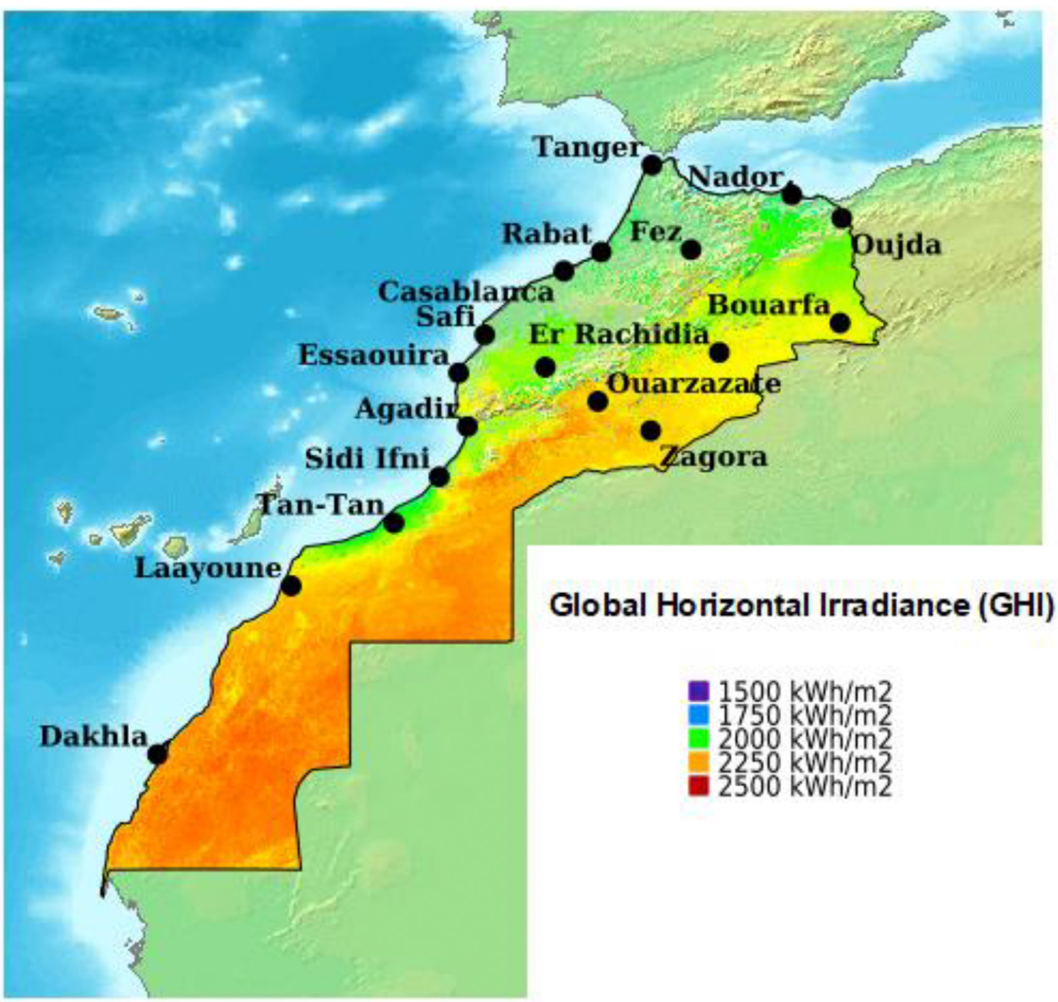

Fig. 2. Global solar energy in Morocco [40].

weather conditions, which will lead to high forecast error due to poor model training [43]. Moreover, not all of the data given are useful. First, all of data is loaded as appropriate data types. The dataset includes columns of date and time. This seems to be the date when the dataset was published with a time step of 30 min. The distribution of solar energy during the period studied is shown in Fig. 3, and the variation of solar energy according to the hours of the day is given in Fig. 4.

It is clear that this distribution depends on each season of the year. The summer and spring months show the highest amount of solar energy as shown in Fig. 5.

Table 1 presents the various values of solar energy production

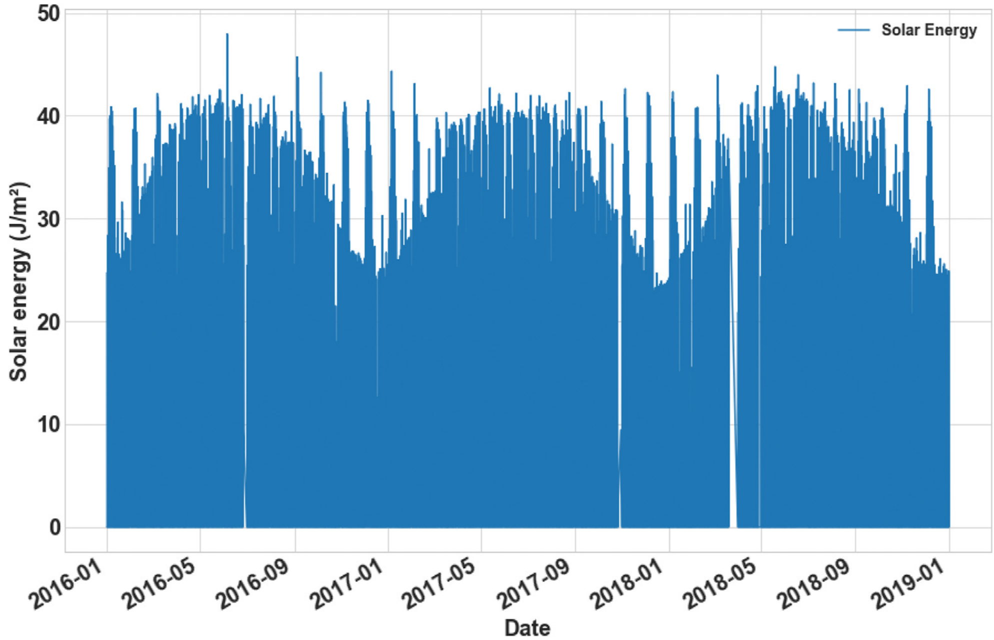

Fig. 3. Distribution of solar energy during the period studied (2016–2018).

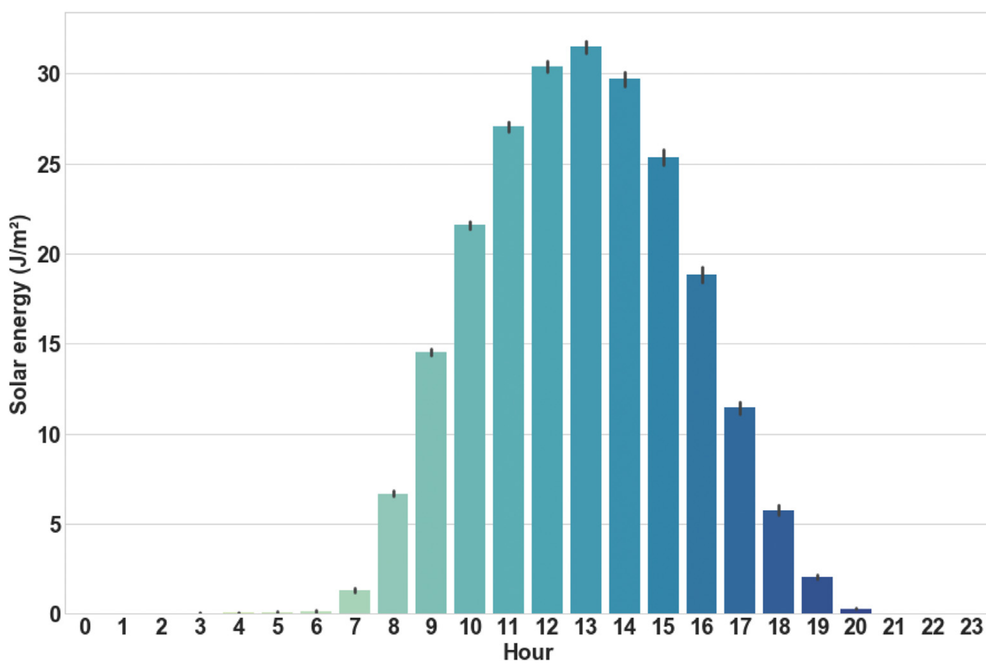

Fig. 4. Solar energy per hour.

during each month of the studied period. We note that the solar energy generation doesn't vary much from year to year in recent years (163000.31 J/m^2, 162608.73 J/m^2, and 161719.69 J/m^2, respectively, in 2016, 2017, and 2018).

2.5. Feature extraction and identification

As mentioned above, solar energy generation is strongly dependent on weather conditions that can affect it to various degrees. Therefore, feature extraction and identification is one of the most important steps in the field of energy forecasting. It is necessary to identify which characteristics selected from the used data and the associated meteorological variables, are containing the most relevant information helping to provide accurate projections [44]. In this stage, we import plotting libraries to visualize data, and we analyze the correlations between solar energy and the other parameters before elaborating the predictive model. For this purpose, we use Pearson correlations to determine which parameters have the biggest impact on each other [45]. As a first step, a basic correlation matrix is generated to remove irrelevant data and identify the most significant characteristics. The Pearson correlation coefficient refers to a measure of the linear dependence

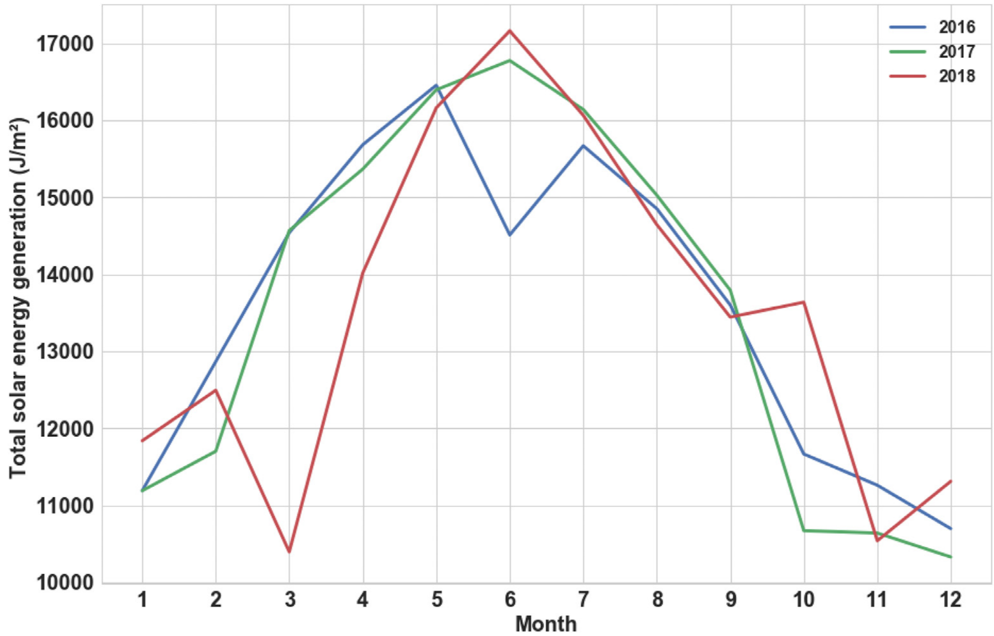

Fig. 5. Cumulative solar energy generation (Monthly).

Table 1
Solar energy production per month of the studied period (2016–2018).

Year	Total solar energy generation (J/m^2)											
	January	February	March	April	May	June	July	August	September	October	November	December
2016	11190.62	12866.82	14540.47	15682.92	16455.68	14511.96	15669.83	14852.98	13601.68	11667.26	11262.14	10697.95
2017	11190.59	11702.96	14568.78	15366.48	16396.21	16776.28	16141.31	15029.09	13793.78	10671.93	10641.17	10330.15
2018	11838.89	12494.69	10396.43	14017.84	16162.45	17161.85	16064.99	14646.05	13444.87	13638.04	10540.74	11312.85

between two random variables (real-valued vectors). Historically, it is the first formal measure of correlation and still the most widely used measure [46]. This linear correlation coefficient, used to reflect the linear correlation between two normal continuous variables, is described as follows [47]:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \sum (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \tag{2}$$

where.

$\bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i$ denotes the mean of x. $\bar{y} = \frac{1}{n} \sum_{i=1}^{N} y_i$ denotes the mean of y and the coefficient r_{xy}.

Eq. (2) varies between −1 and 1 and it is invariant to the linear transformations of the two variables.

- $r_{xy} = 1$, x and y are a totally positive correlation,
- $r_{xy} = 0$, the linear correlation between x and y is not obvious,
- $r_{xy} = -1$, x and y are a totally negative correlation.

After exploring the data, we observe that in dataset:

- Higher temperatures are linked to a greater amount of solar energy. This is confirmed by a Pearson R-value of 1 and by the observed high solar radiation output following the temperature on the daily and weekly time scales.
- Humidity has a smaller impact on the amount of solar energy with a Pearson R-value (−0.36).
- Humidity cannot be ignored since it is a potential system factor.
- Pressure doesn't correspond much to solar energy with a Pearson R-value (−0.12), but it is consistent with the temperature and humidity. As temperature, pressure, and humidity are all characteristics of the atmosphere, these characteristics are correlated.
- Wind speed and direction are not pertinent in this analysis. Although both are characteristic of the local climate, they make no sense in predicting solar energy.
- Wind direction has a moderate correlation with temperature (0.33), pressure (−0.42), solar radiation, and solar energy (0.23). This is just a correlation and not causality. The correlation values are presented by the Pearson matrix in Fig. 6.

Weekly time scales are the best predictors. The variation in solar energy from month to month is important to account for seasonal changes within the same year. Daily and hourly measurements generate a lot of noise when we are looking for seasonal changes. Because temperature, pressure, and humidity dominate daily, the week of the year is the best indicator of seasonal trends. Hence, also although it is clear, solar radiation and solar energy have a strong correlation with time of day. Thus, we add the time of day as a feature so that the algorithm can differentiate between day and night. This implicitly explains the angle of the sun.

Based on the most correlated characteristics, solar energy is plotted as a function of solar radiation, temperature, humidity, and pressure on different time scales. Fig. 7 shows the strong correlation between solar energy and solar radiation as a function of their hourly and weekly averages. Furthermore, Fig. 8 shows the strong

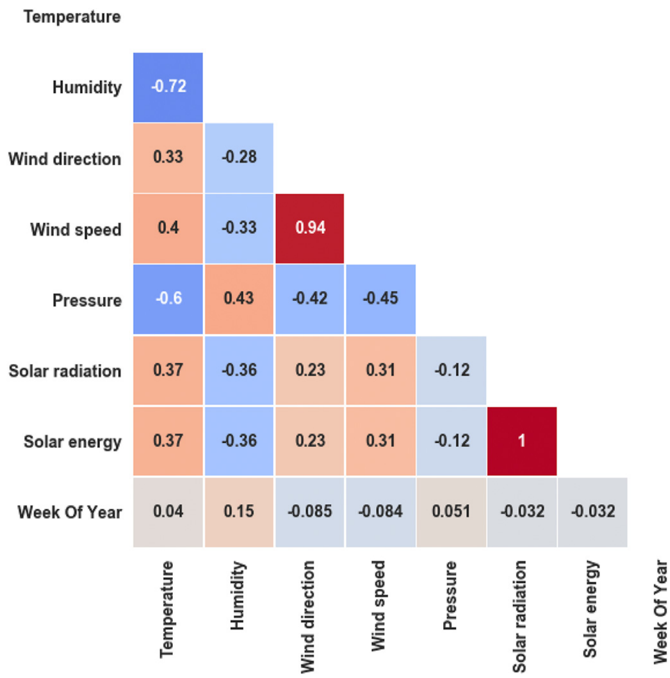

Fig. 6. Pearson correlation of the solar energy parameters.

correlation between solar energy and temperature as a function of their hourly and weekly averages. However, the negative correlation between solar energy and humidity as a function of their hourly and weekly averages is shown in Fig. 9. Besides, Fig. 10 shows the negative correlation between solar energy and pressure as a function of their hourly and weekly averages.

3. Learning methodology

The adopted methodology in our study consists of learning from historical and the most relevant data selected after using Pearson correlation. This section describes the used learning models, namely LR, RF, SVR, and ANN, and also the performance metrics. Note that we have implemented the models using Python libraries: Scikit-learn, Pandas, NumPy, SciPy, and Matplotlib in a Jupiter notebook. A train/split test methodology was also adopted while training models, to avoid learning bias. The dataset was split into a group of randomly sampled data points. 80% of these data were used for training and the remaining 20% were used for testing and validation.

3.1. Studied models overview

Linear regression is a very basic supervised learning algorithm that aims to find the best possible line using a single explanatory variable. It is a univariate model, as opposed to multivariate models, which use several variables [48]. In the case of linear regression, the hypothesis function h takes a more general form for n input variables as given in the following formula:

$$h(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \ldots \theta_n X_n \tag{3}$$

where $\theta(x)$ is the feature vector of inputs x.

RF [49] is a classification and regression learning algorithm which trains numerous simple decision trees and groups the results of each tree. Using random subsets of the dataset, each one is trained to recognize different patterns and synergies in data. The algorithm usually consists of three major steps as described below [50,51]: (i) Creating B sample datasets of size N from the training data, (ii) generating an RF tree $T_b$for each sample dataset, by repeating the following steps for each terminal node, until the minimum node size n_{min} is achieved, and (iii) finding the sets of the trees $\{T_b\}_1^B$, where B is the number of trees in the RF. The forecast

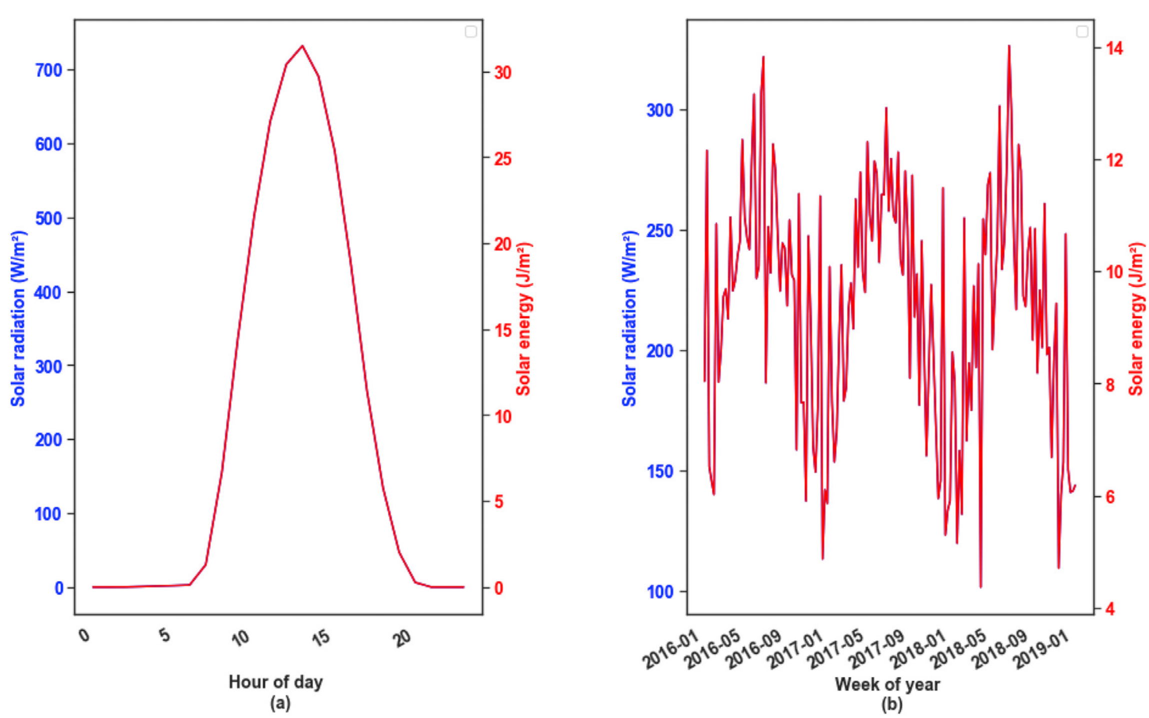

Fig. 7. Variation of the average solar radiation and average solar energy; (a) Hourly and (b) Weekly.

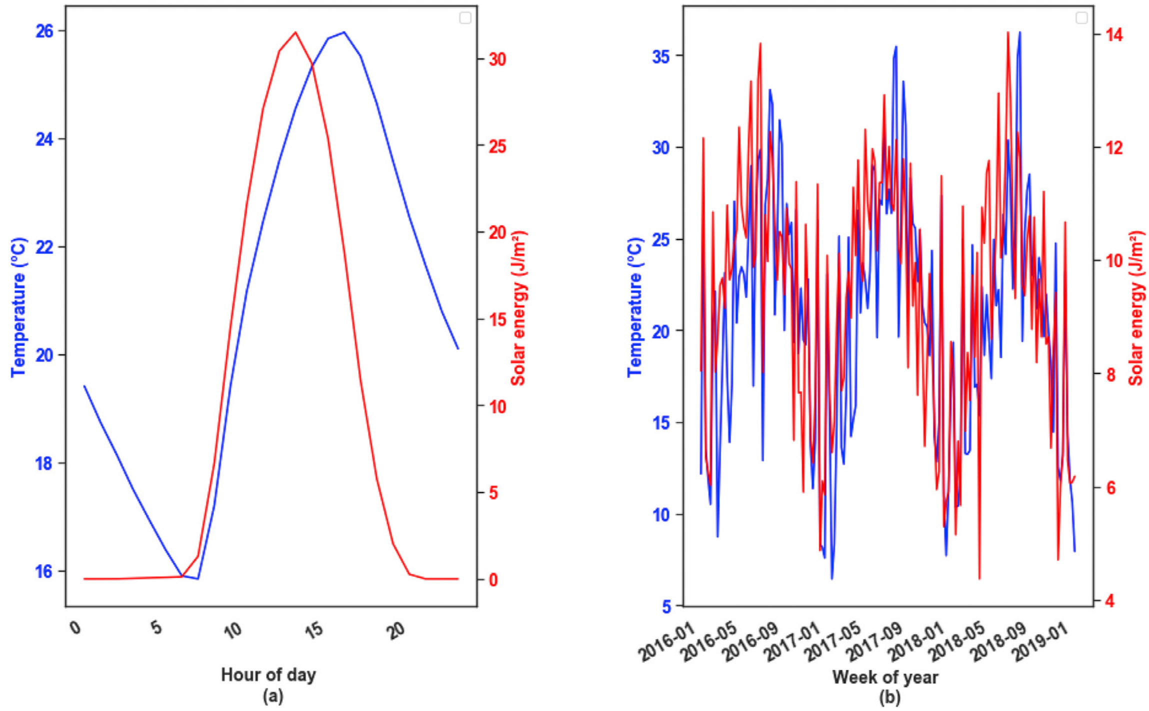

Fig. 8. Variation of the average temperature and average solar energy; (a) Hourly and (b) Weekly.

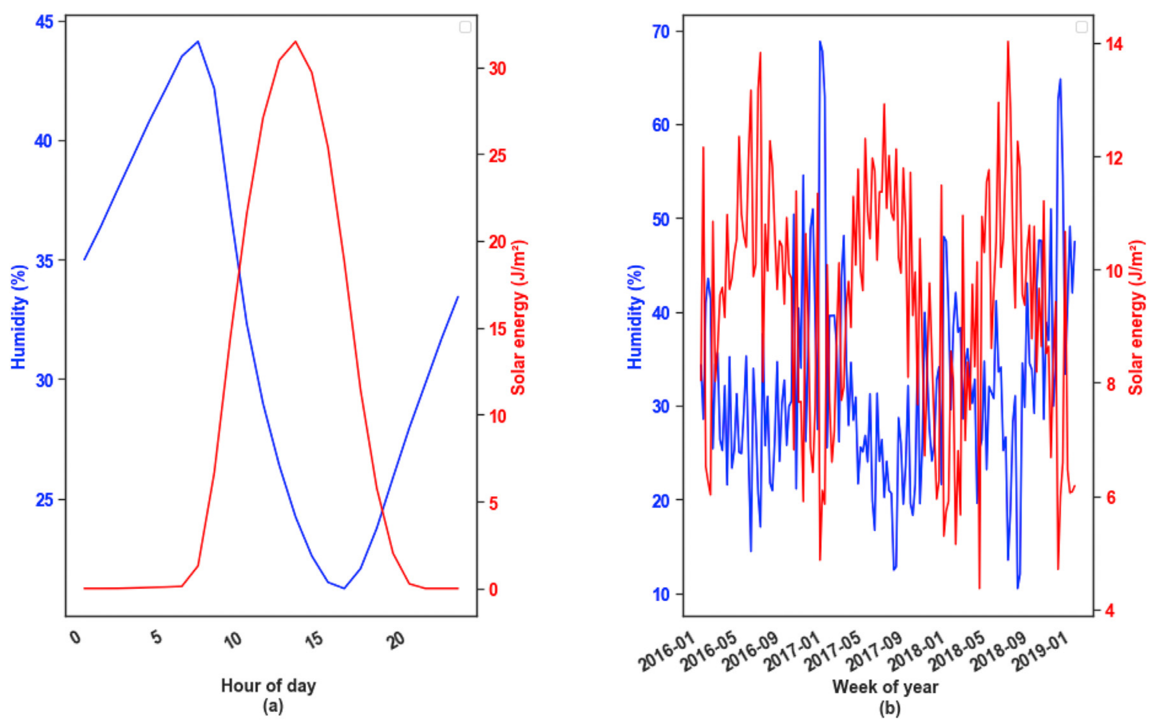

Fig. 9. Variation of the average humidity and average solar energy; (a) Hourly and (b) Weekly.

is the mean of the different regression tree results, and the RF algorithm is used to build the f presented in Eq. (4):

$$\widehat{f}_{RF} = \frac{1}{B}\sum_{b=1}^{B} T_b(x) \quad (4)$$

In our study, we have considered the following parameters: number of trees B = 100, samples m = 2, leaf size $n_{min} = 1$.

SVM is a supervised ML method that makes data correlation by non-linear mapping. It is a direct method of computing a kernel function [18], and its implementation includes two steps [38]: Mapping learning points by a non-linear function to a large dimensional space in which the points are linearly divisible, and determining the optimal dividing hyperplane that maximizes the

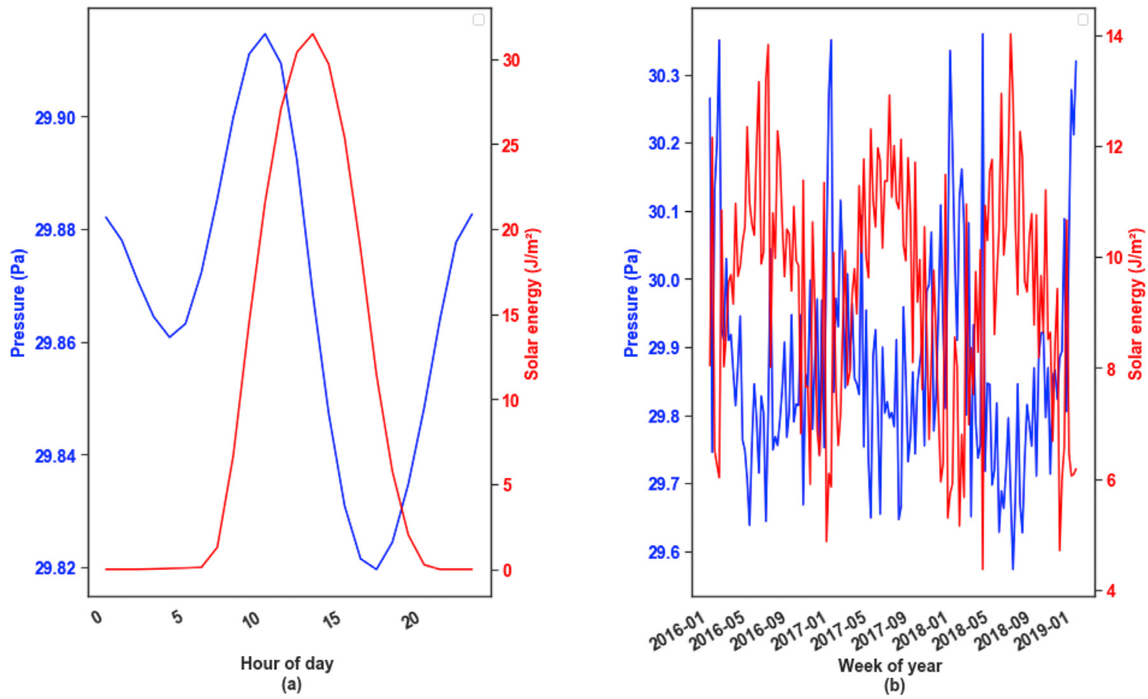

Fig. 10. Variation of the average pressure and average solar energy; (a) Hourly and (b) Weekly.

distance between the points of each category. SVR which a is non-linear regression algorithm, consists of a non-linear mapping of the input time series data sample into a larger function space and then executing a linear regression in this space [5]. It takes into account $\{x_i, y_j\}_{i=1}{}^N$ a training set, where $x_i \in R^n$ is the input vector (meteorological variables data) and $y_i \in R$ is the corresponding output value (solar energy). As given in Ref. [17], the used function $f(x)$ is shown below:

$$y_f = f(x) = w \times \phi(x) + b \tag{5}$$

where $\phi(x)$ is the feature vector of inputs x, $w \in R^n$ is a weight vector and $b \in R$ is the bias term. We note that in this study, we have used the default Radial Base Function and default parameters (C = 1 and Gamma = 'scale') while using SVR.

ANN [52] is a very powerful computational tool for complex problems, which shows adaptive behavior for complex and noisy information. Thanks to its ability to work with nonlinear relationships, it is based on known input and output values. The algorithm modifies the weights of the hidden and output nodes (neurons) until the output approaches the real data within the limits of a given threshold [53–55].

The mathematical expression of the output neuron is defined by Eq. (6).

$$A_i = g\left(\sum_{i=0}^{n} W_{ji} * aj\right) \tag{6}$$

where A_i is the output of network, W_{ij} is the connection weight of jth neuron to ith layer neuron, and aj is the input of the neuron. Note that in our study, we used Multilayer Perceptron (MLP) which is a network of interconnected nodes of multiple layers. We note that to perform predictions with this model, we have used ReLU, and finally tanh as the activation function, and opted for ADAM which is the most popular powerful optimizer thanks to its ability to make the model learns fast. We have also experimented with the model with two layers containing 100 neurons, and finally, opted for eight layers containing, respectively, 100, 200, 300, 50, 25, 10, 5, and 1 neurons with a batch size equal to 1.

3.2. Performance metrics

In general, the performance of a model is evaluated by computing the deviation between predicted values and actual observations using different statistical methods [56]. For this purpose, several metrics are used to evaluate ML model results including the Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), Max Error (ME), and R-squared (R2). Their formulas are given below:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|G_i - GP_i| \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(G_i - GP_i)^2 \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(G_i - GP_i)^2} \tag{9}$$

$$ME = \underset{1\le j\le n}{Max}\,|G_i - GP_i| \tag{10}$$

$$R_{Squared} = 1 - \frac{SS_{regression}}{SS_{total}} \tag{11}$$

where G is the actual output, GP is the expected output and n is the number of samples. Eq. (7) calculates MAE as the average of the absolute errors (absolute values corresponding to the differences between the actual and the predicted values). Eq. (8) measures MSE that is the average squared errors (the difference between the real

Table 2
Metrics values related to Moroccan (Errachidia) scenario 1.

Methods	MAE (J/m^2)	MSE (J/m^2)	RMSE (J/m^2)	Max Error (J/m^2)	R-squared (%)
LR	0.0013	4.36e-06	0.002	0.005	99.99
RF	2.64e-05	9.93e-07	0.0009	0.09	99.99
SVR	0.04	0.005	0.07	3.93	99.99
MLP (ANN)	0.03	0.006	0.08	6.52	99.99

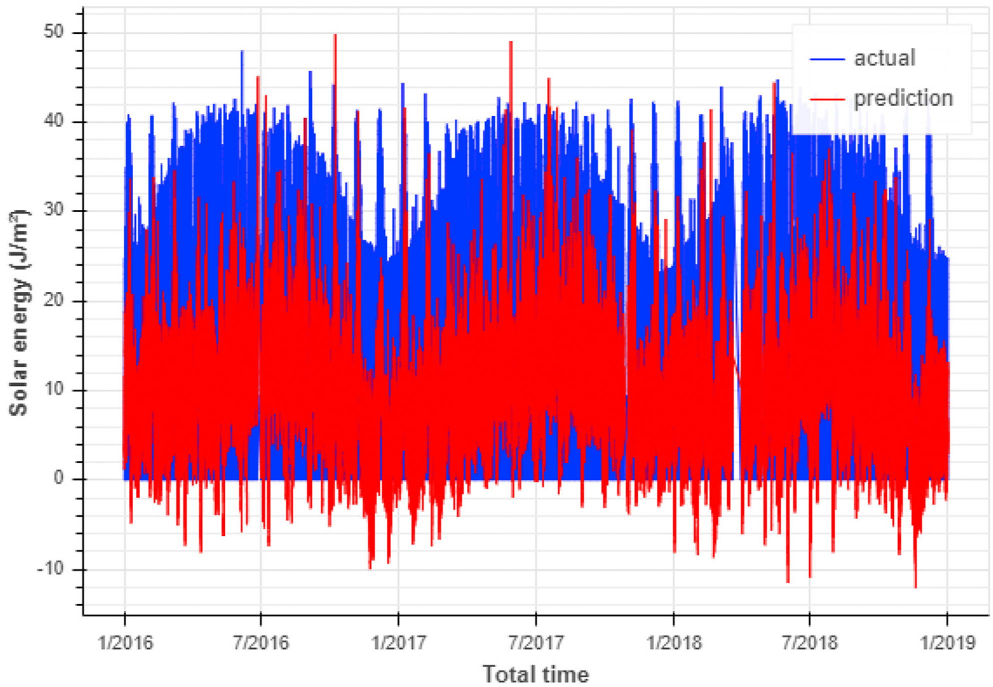

Fig. 11. Real and predicted values of solar energy using LR (Morocco, Errachidia).

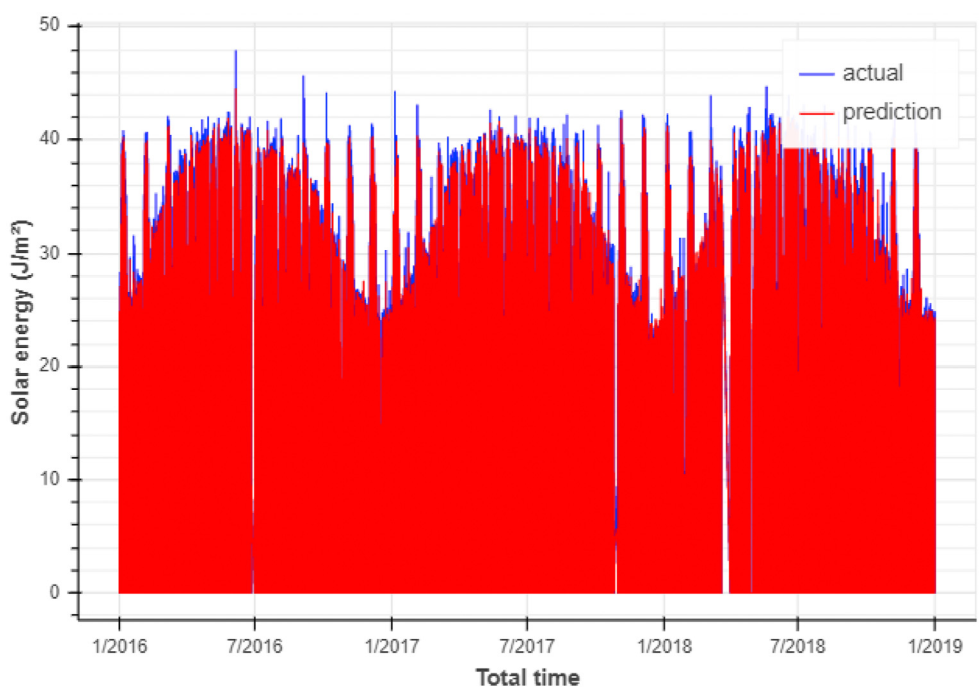

Fig. 12. Real and predicted values of solar energy using RF (Morocco, Errachidia).

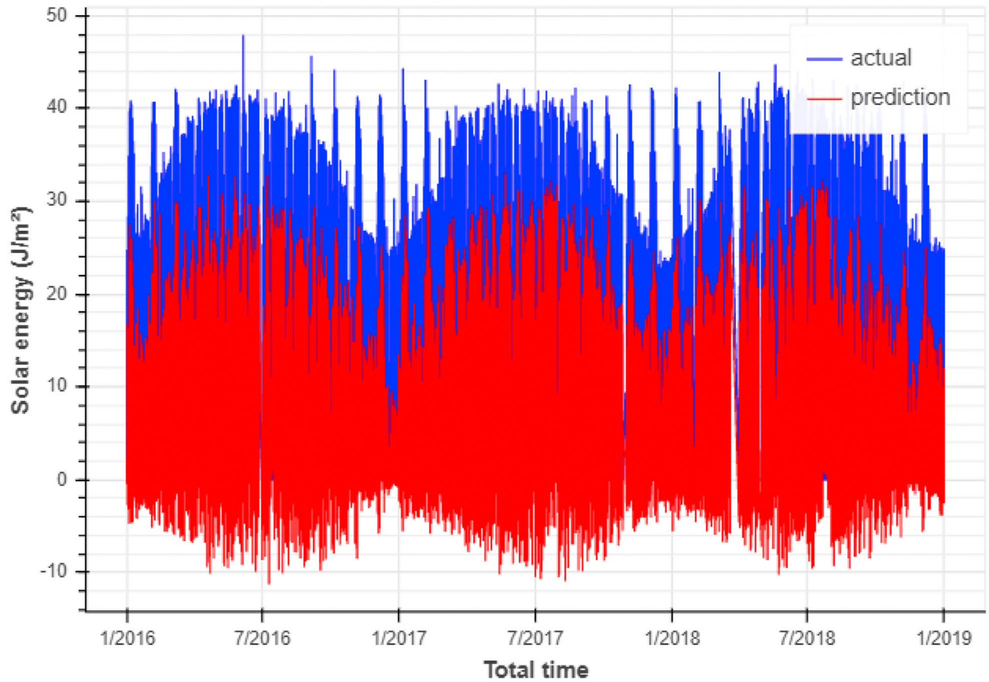

Fig. 13. Real and predicted values of solar energy using SVR (Morocco, Errachidia).

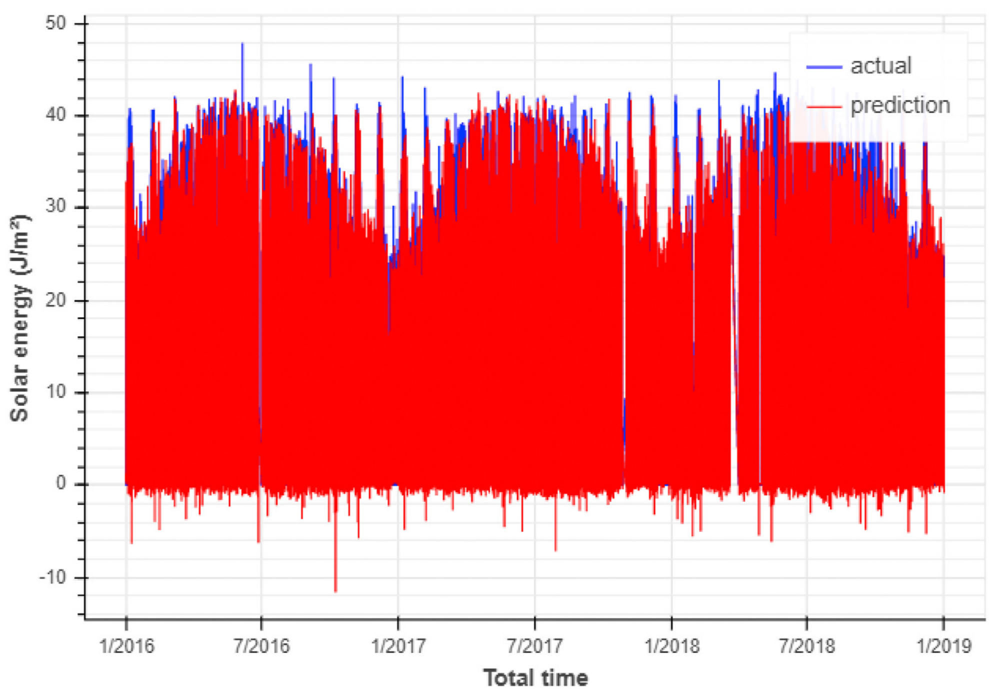

Fig. 14. Real and predicted values of solar energy using ANN (Morocco, Errachidia).

values *G* and what is estimated *GP*). Eq. (9) calculates RMSE which is the square root of the MSE, and is used when only small errors are tolerated [57], whereas Eq. (10) measures the maximum residual error ME and highlights the worst error between the real and the predicted values. Besides, in order to calculate the predictive accuracy of our models, we also used R-squared, which represents a statistical measure in a regression model, to determine the proportion of variance in the dependent variable that can be made explicit by the independent variable. In other words, Eq. (11) shows the measure R-squared to which the data fit the regression model (the goodness of fit) where $SS_{regression}$ is the sum of squares due to regression and SS_{total} is the total sum of squares [58].

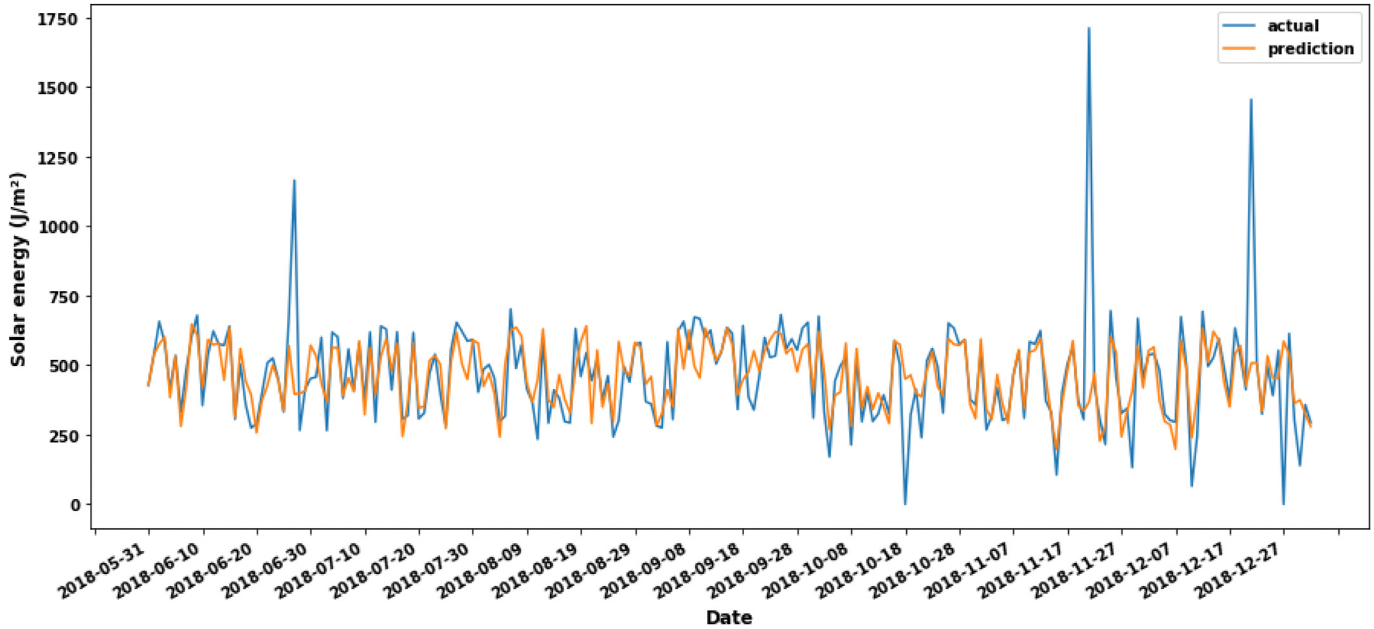

Fig. 15. Real and predicted values of solar energy using LR for daily prediction (Morocco, Errachidia).

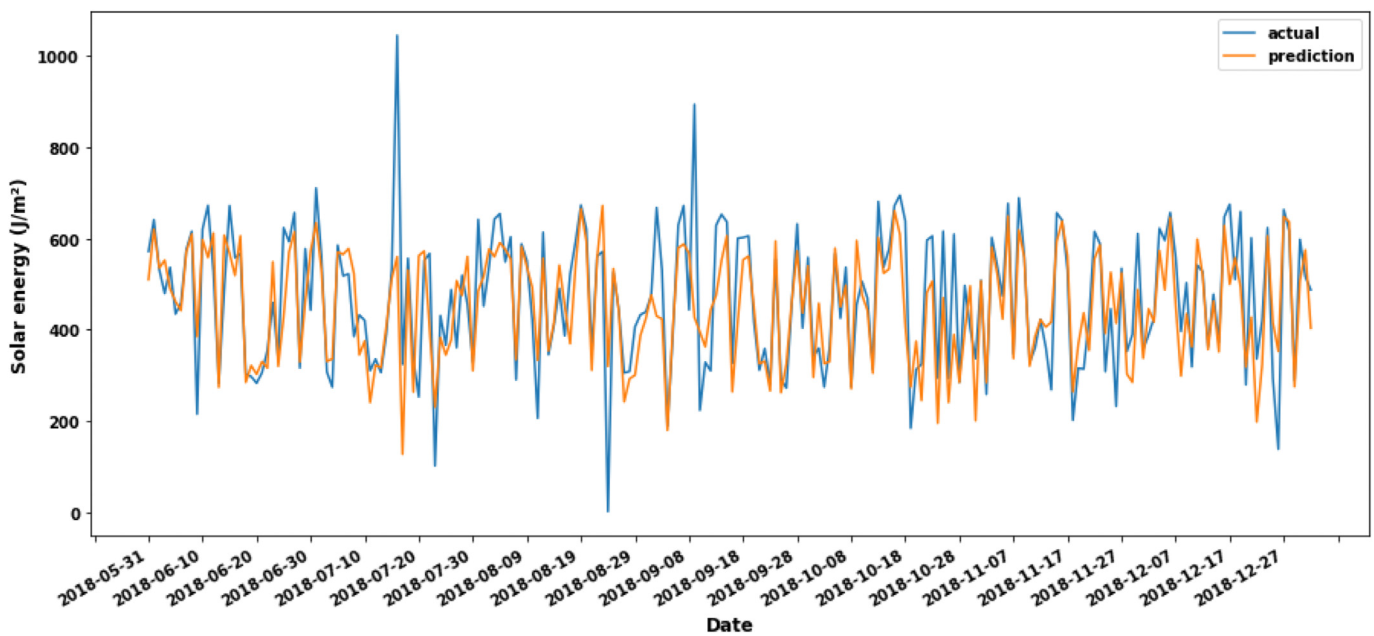

Fig. 16. Real and predicted values of solar energy using RF for daily prediction (Morocco, Errachidia).

4. Results and discussion

4.1. Results and key performance indicators

In this section, we present the results of our four studied models tested for solar energy forecasting while considering tree scenarios. Scenario 1 consists of learning from all available data without applying the Pearson correlation. Scenarios 2 and 3 consist of learning from only the most correlated data to perform respectively, real-time (predicting solar energy for the next half-hour) and daily solar energy forecasts. After presenting the obtained results, we discuss them to highlight the key performance indicators (KPI) of our approach, and especially of the adopted ANN model.

As illustrated in Table 2 given in the next section, the R-squared accuracy related to scenario 1, is equal to 99.99% for all of the elaborated models. This too high value of R-squared accuracy is the consequence of the model's complexity, and it depends on data rather than the studied algorithms. In other terms, it is one of an overfit model's symptoms. Using variables that aren't the most relevant ones, make a statistical model unable to describe the relationships between variables, and lead to an overfitting issue where the model begins to describe the random error in the data. This is why we use the Pearson correlation coefficient in scenario 2 and 3, to select the relevant and most correlated meteorological features, and thus avoid the overfitting issue as confirmed by the obtained results.

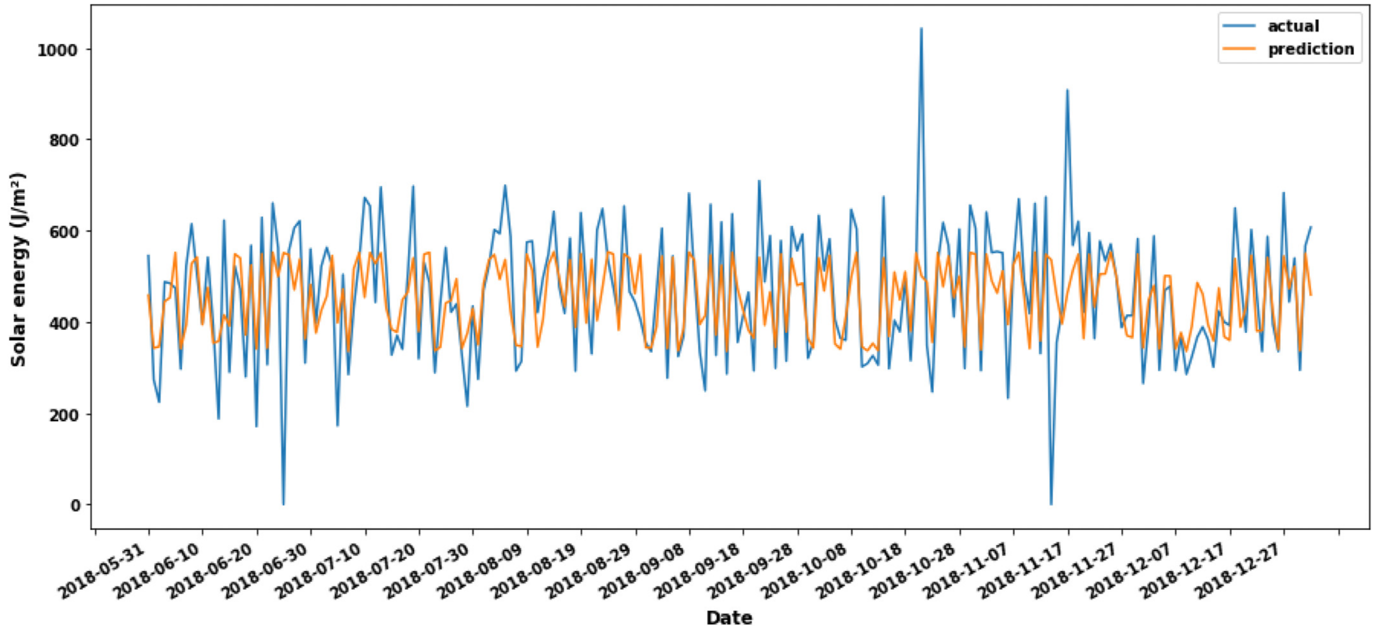

Fig. 17. Real and predicted values of solar energy using SVR for daily prediction (Morocco, Errachidia).

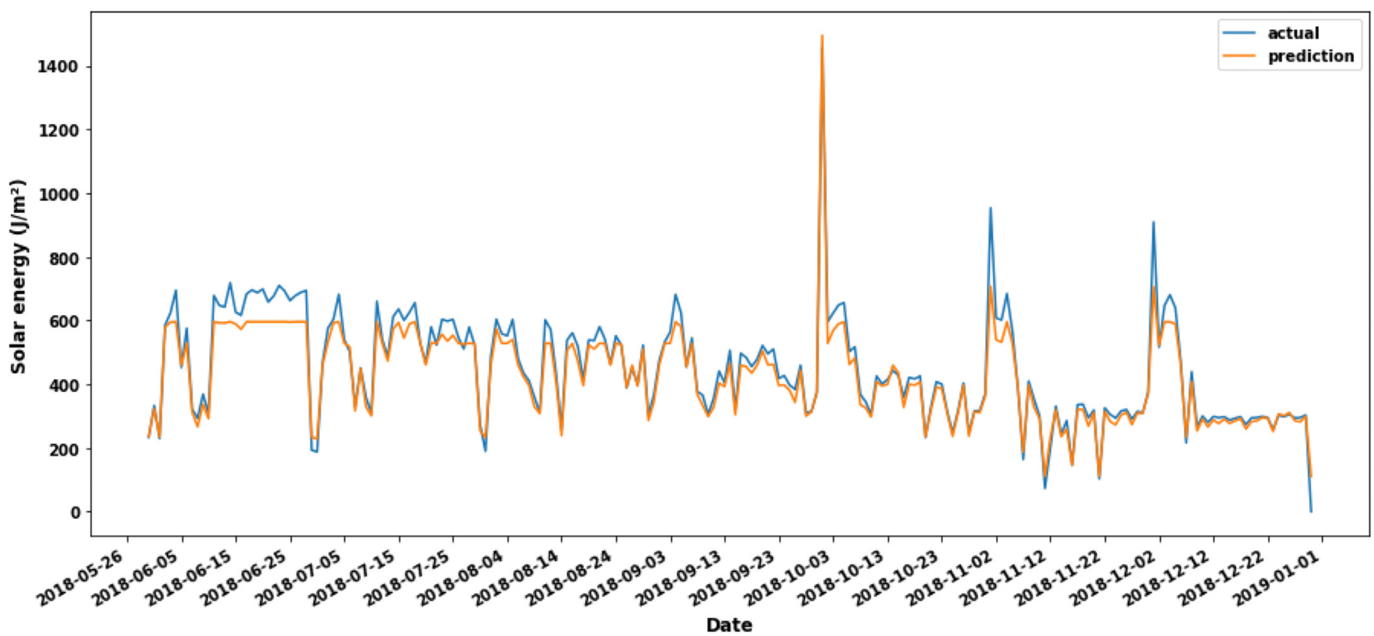

Fig. 18. Real and predicted values of solar energy using ANN for daily prediction (Morocco, Errachidia).

Table 3
Metrics values related to Moroccan (Errachidia) scenario 2.

Methods	MAE (J/m^2)	MSE (J/m^2)	RMSE (J/m^2)	Max Error (J/m^2)	R-squared (%)
LR	9.30	128.72	11.34	41.77	22.05
RF	1.47	8.45	2.90	25.97	94.81
SVR	5.92	61.05	7.81	35.10	62.62
MLP (ANN)	1.94	12.19	3.49	29.40	92.61

Fig. 11, Fig. 12, Fig. 13, and Fig. 14 show the curves corresponding to real and predicted values of solar energy using respectively, LR, RF, SVR, and ANN while considering scenario 2, whereas, Fig. 15, Fig. 16, Fig. 17, and Fig. 18 show the same curves corresponding to the scenario 3. To sum up, according to the R-squared metric, the ANN and the RF models give better results for real-time prediction with respectively, a very good accuracy of 92.61% and 94.81% as shown in Table 3, whereas, for daily prediction, ANN outperforms

Table 4
Metrics values related to Moroccan (Errachidia) scenario 3.

Methods	MAE (J/m^2)	MSE (J/m^2)	RMSE (J/m^2)	Max Error (J/m^2)	R-squared (%)
LR	79.2	24292.99	155.86	1345.74	31
RF	67.33	9399.16	96.94	483.5	58
SVR	77.74	11919.53	109.17	551.13	44
MLP (ANN)	30.3	1980.24	**44.49**	245.72	**93**

Table 5
Results synthesis for real-time and daily Moroccan (Errachidia) predictions.

Model	Prediction	NRMSE	Training time (s)
LR	**Real-time**	0.25	0.012
	Daily	0.14	0.0026
RF	**Real-time**	0.06	12.88
	Daily	0.092	0.35
SVR	**Real-time**	0.17	79.58
	Daily	0.10	0.034
MLP	**Real-time**	0.07	45.52
	Daily	**0.03**	124.14

the three other models with very good accuracy of 93%.

In order to measure the performance and quality of these results, we evaluate the studied models regarding a set of criteria that also constitute the KPI of our approach and adopted model ANN namely, (i) adaptation to solar energy context regarding the ability to support non-linear models, (ii) performance and results in quality regarding the predictive accuracy, (iii) reliability, (iv) ability to perform deep learning from huge data and then to be adapted to especially, long-term prediction, and (v) ability to perform fast learning in terms of training speed and calculation cost. Our conclusions are then based on the results of our work and also the characteristics of the used learning models.

The datasets used in solar energy prediction, are characterized by non-linearity and complexity. They have a non-stationary and random process, which is influenced by weather conditions due to the chaotic nature of the weather. Non-linearity is supported by RF [59]. However, according to the literature, ANNs and SVRs are good tools when there is a complex and non-linear link between data without a prior hypothesis. In addition, the performance of a predictive model is highly dependent on the correlation between their input and output values [60]. This is why our approach and all of the studied models use the Pearson correlation coefficient. This one was helpful to identify the highly correlated meteorological features which are relevant to be used as input variables for accurate solar energy forecasting. However, in terms of empirical results, ANN has provided accurate predictions in both real-time and daily forecast cases. In terms of reliability, according to the literature, ANNs have high self-adaptation, fault tolerance, robustness, and deductive capabilities, provide a strong correlation between measured data and predicted values, and exceed conventional mathematical models in terms of accuracy and adaptability [61]. The proposed ANN model can then be considered a reliable and efficient model. Moreover, ensuring efficiency and providing accurate projections, also involves the ability to process large size of input data to better learn from more historical information. We note that in the literature, both SVR and ANN models are suitable for learning from large datasets. Hence, our ANN model which shows very good and promising accuracy results, still suitable to be also experimented with and used for both long-term and short-term solar energy forecasting from huger datasets.

Finally, although ANN like all other deep learning models, performs deep processing to learn, and then could lead to more computing complexity and be a high-cost solution in terms of time and resource consumption, it remains a promising and suitable solution to be adopted for solar energy forecasting. The use of deep learning models, among them the ANN one, is nowadays encouraged thanks to the technological progress in terms of both fast processors and advanced storage tools and techniques like cloud computing. Moreover, during our empirical experimentations, we observed that SVR have a high training speed compared to ANN and RF and that these ones which are extensively used in practice due to their ability to solve complex and non-linear prediction models have shown a good training speed (respectively 12.88 s and 79.58 s were reported by RF and ANN while performing real-time forecasts from a dataset of 4,93 mega octet, against 0.034 s and 124.14 s for daily predictions and a dataset of 95,3 kilo octet). This still matches with the norms in the practice. In addition, note that the time necessary to perform the training phase and then to elaborate the predictive model doesn't compromise the subsequent use of the concerned model. Once this one is elaborated, it can then be used to give solar energy forecasts in very fast times. Besides, advanced recent techniques such as Resnets [62] are nowadays to reduce the

Table 6
Metrics values related to hourly predictions in Brazil (Pirapora).

Methods	MAE (KJ/m^2)	MSE (KJ/m^2)	RMSE (KJ/m^2)	Max Error (KJ/m^2)	R-squared (%)
LR	658.16	751881.46	867.11	2609.07	46.95
RF	130.42	72141.58	268.59	2572.40	**94.90**
SVR	841.22	2040190.82	1428.35	4063.81	−45.17
MLP (ANN)	299.88	159134.63	398.91	2203.89	**91.53**

Table 7
Metrics values related to daily predictions in Brazil (Pirapora).

Methods	MAE (KJ/m^2)	MSE (KJ/m^2)	RMSE (KJ/m^2)	Max Error (KJ/m^2)	R-squared (%)
LR	3679.72	20881422.29	4569.61	10850.76	39
RF	2673.33	13912522.55	3729.94	18794.62	**52**
SVR	4022.84	25431265.63	5042.94	13625.65	−1
MLP (ANN)	891.23	1626501.32	1275.34	4767.11	**95**

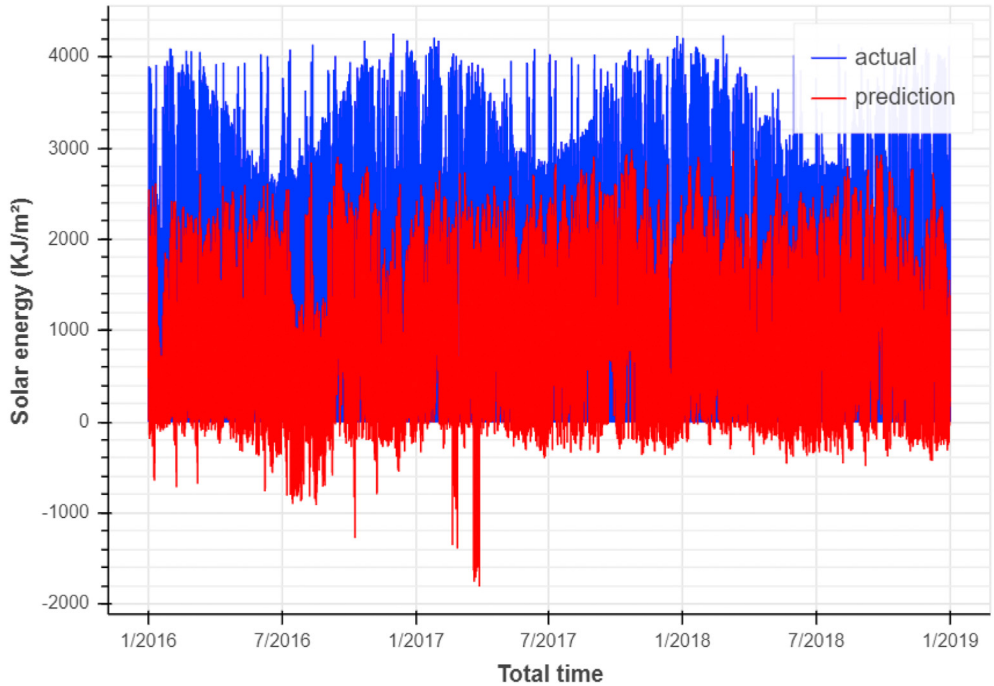

Fig. 19. Real and predicted values of solar energy using LR (Brazil, Pirapora).

complexity of deep learning algorithms and ensure both efficiency and low-cost processing. All these observations lead us to conclude that among the models studied, ANN is better suited for making solar energy predictions.

4.2. Analysis of the statistical metrics

It is worth noting that although R-squared (R2) is a widely used metric, it isn't considered an appropriate indicator of the model's fit

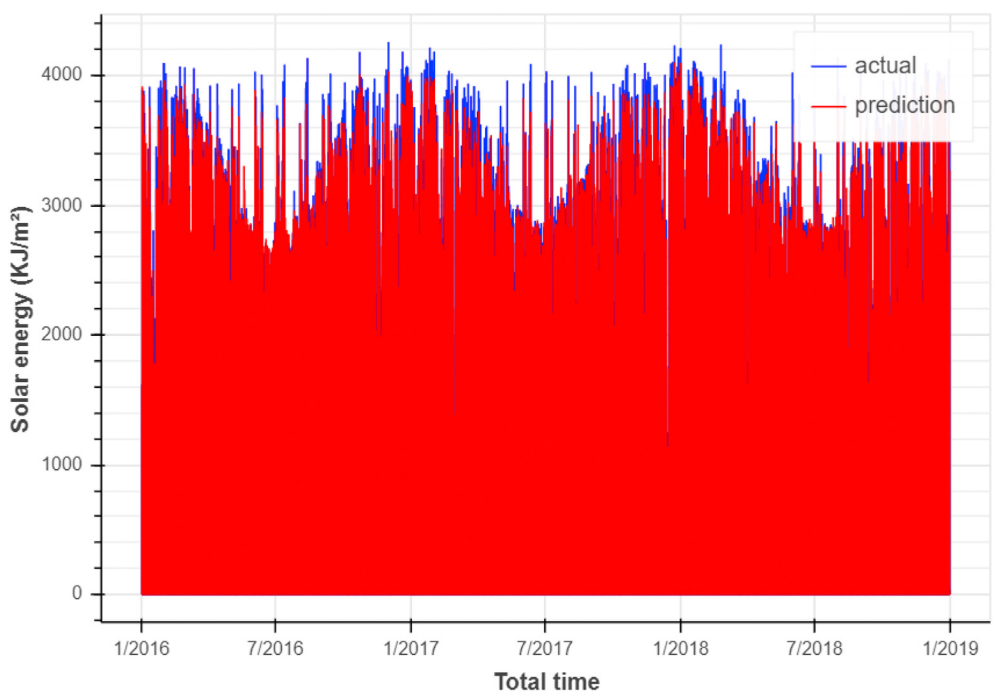

Fig. 20. Real and predicted values of solar energy using RF (Brazil, Pirapora).

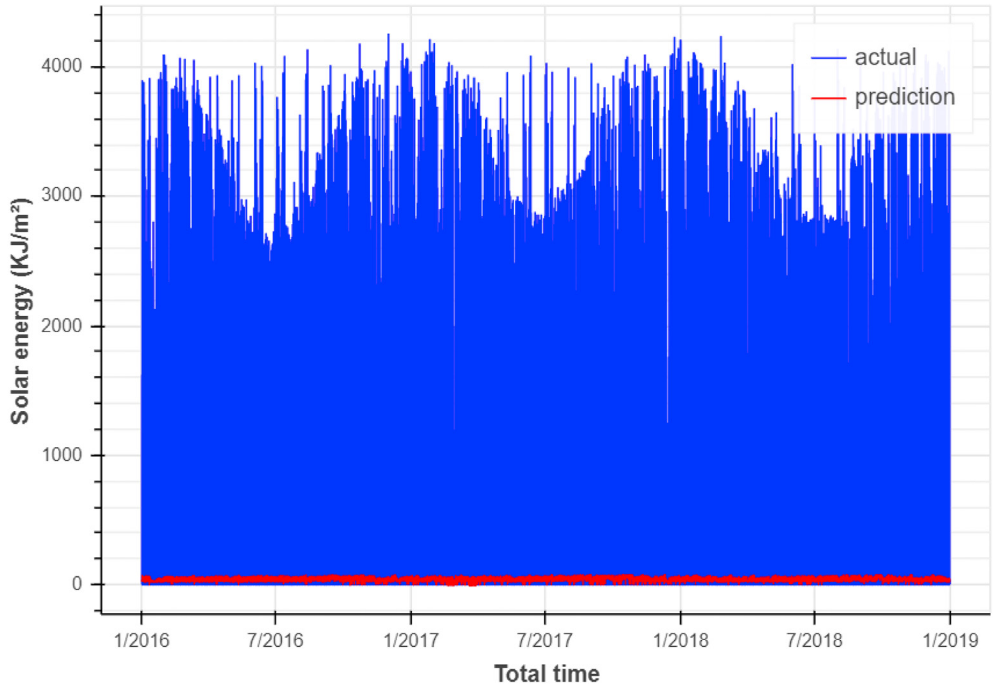

Fig. 21. Real and predicted values of solar energy using SVR (Brazil, Pirapora).

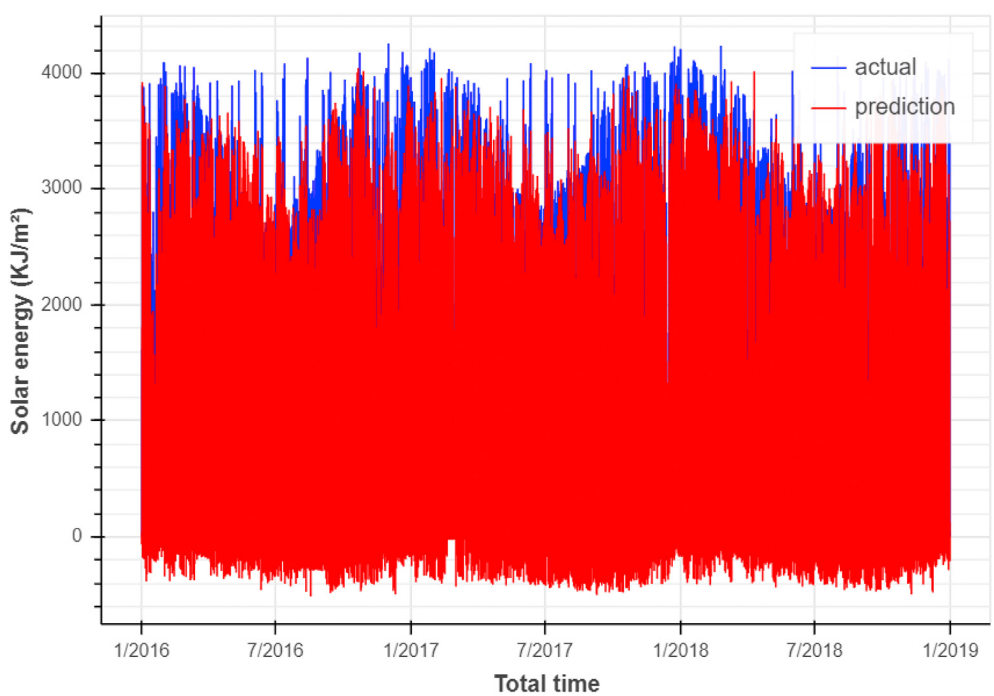

Fig. 22. Real and predicted values of solar energy using MLP (Brazil, Pirapora).

with the data [63]. A small or a high value of R2 does not necessarily indicate that the model is bad or that it is automatically correct [64]. Given the controversy surrounding the efficiency of R2 as an appropriate measure for identifying the best regression model [58,59], we also evaluate our results using other statistical metrics that are the most popular ones: MAE, MSE, Max Error, and RMSE. This last one is still the most commonly used metric in the regression field [8,38,55,65]. In this section, we present and discuss the metrics values related to each studied scenario and each model.

Table 2, **3**, and **4** illustrate the metrics values corresponding respectively to scenarios 1, 2, and 3.

The metrics values in Table 2 confirm the overfitting issue. Besides, according to the results in Table 3 corresponding to solar energy prediction in real-time, we observe that thanks to their ability to solve complex and non-linear forecast models, both RF and ANN perform better with very high accuracy as shown by their related R-squared values which are respectively equal to 94.81% and 92.61%. SVR and LR give precisions less than RF and ANN with

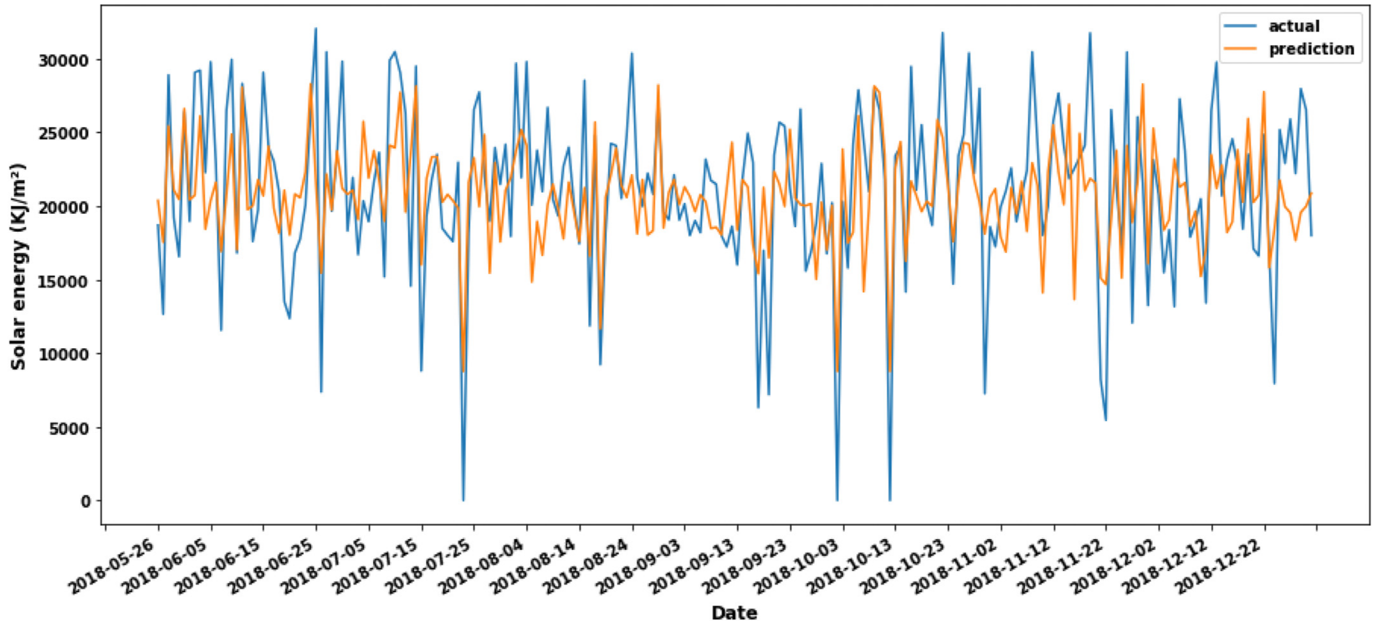

Fig. 23. Real and predicted values of daily solar energy using LR (Brazil, Pirapora).

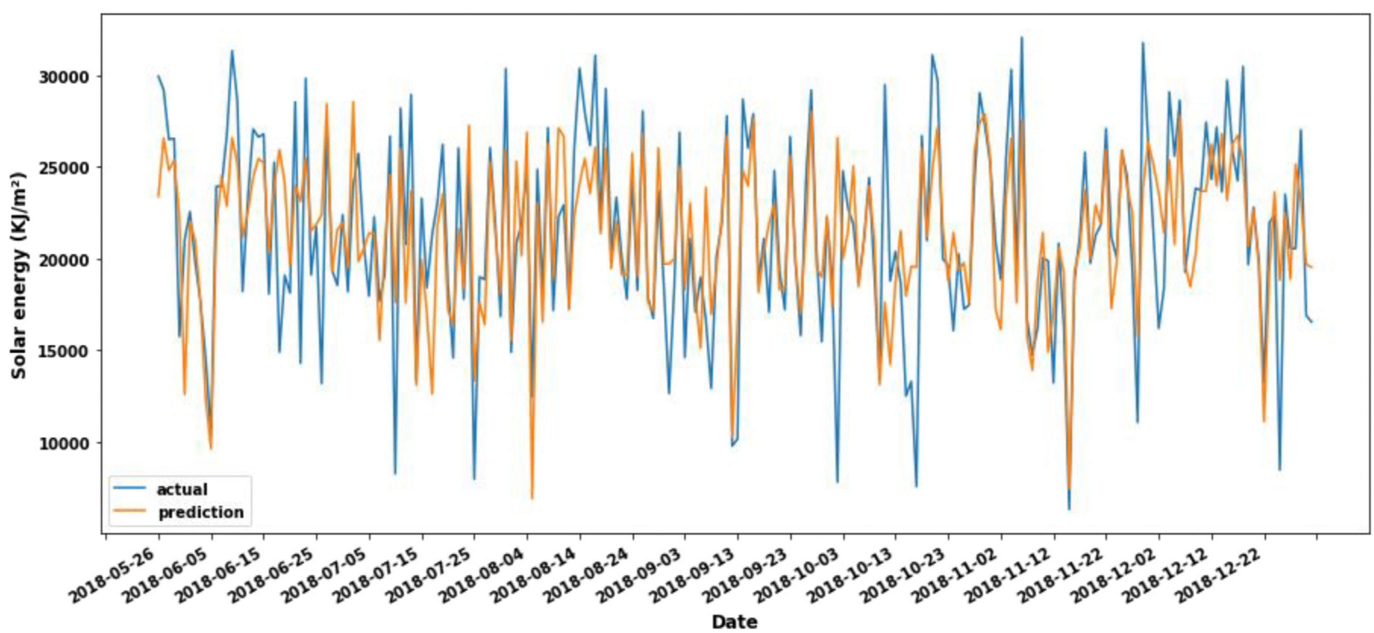

Fig. 24. Real and predicted values of daily solar energy using RF (Brazil, Pirapora).

respectively 62.62% and 22.05% as values. They also lead to very high errors. We note that for RF and ANN, the metrics values are almost similar, in particular for RF we have MAE = 1.47, MSE = 8.45, RMSE = 2.90 and Max Error = 25.97. We can conclude that SVR and LR give poor results, whereas, RF and ANN are showing superior performance. This conclusion can be explained by the fact that SVR and LR are sensitive to huge data while RF and ANN remain robust even when the dataset is very large. In addition, the worst results were reported by LR due to its inability to process complex and non-linear data. To sum up, the results for real-time forecasting show that RF and ANN are both robust and able to minimize errors during learning and test processes.

Besides, when analyzing the results reported in Table 4, we can see that ANN outperforms all the other algorithms with very high accuracy (R-squared = 93%) and minimal errors (MAE = 30.3, MSE = 1980.24, RMSE = 44.49 and Max Error = 245.72). Note that in this case corresponding to daily solar energy forecasting, RF reports poor results comparing to ANN with poor accuracy (R-squared = 58%) and high errors (MAE = 67.33, MSE = 9399.16, RMSE = 96.94 and Max Error = 483.5).

In addition, although RMSE is considered in the literature as a good measure to evaluate and compare regression models, it is still difficult to be well interpreted. Hence, we also normalize the RMSE values to provide a more significant representation of our results which would help to make more reliable conclusions. The normalized RMSE (NRMSE) is the rate of the RMSE value and the

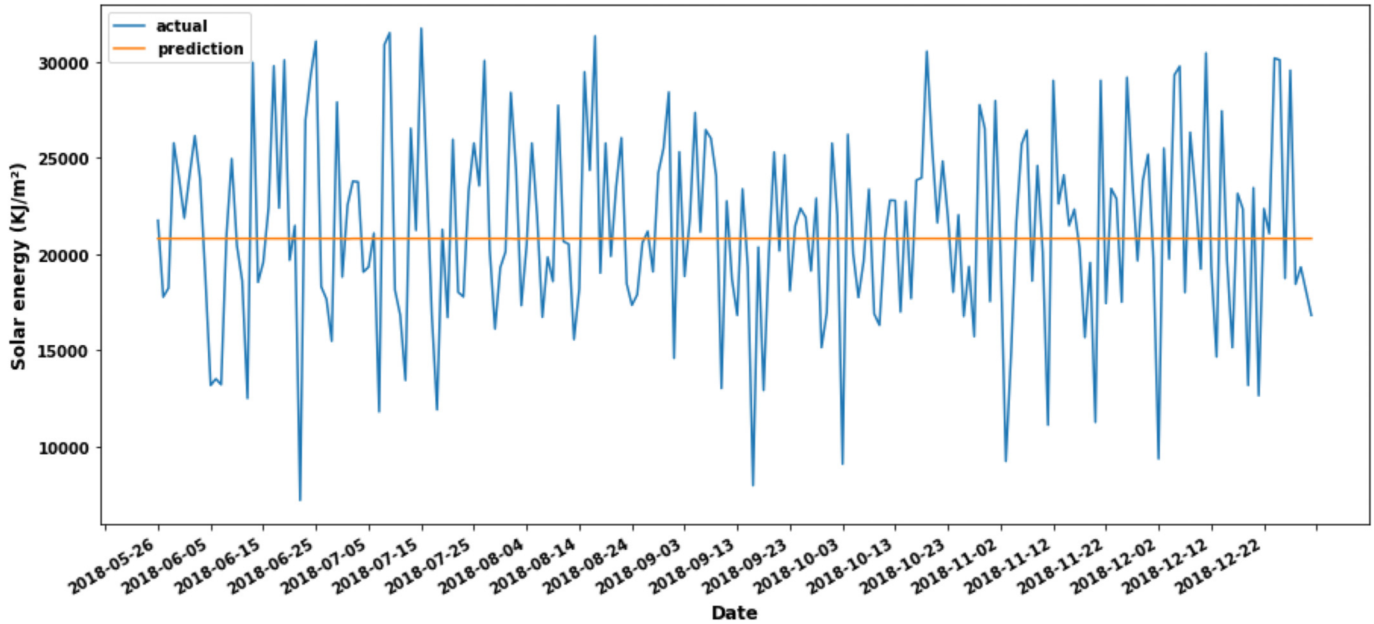

Fig. 25. Real and predicted values of daily solar energy using SVR (Brazil, Pirapora).

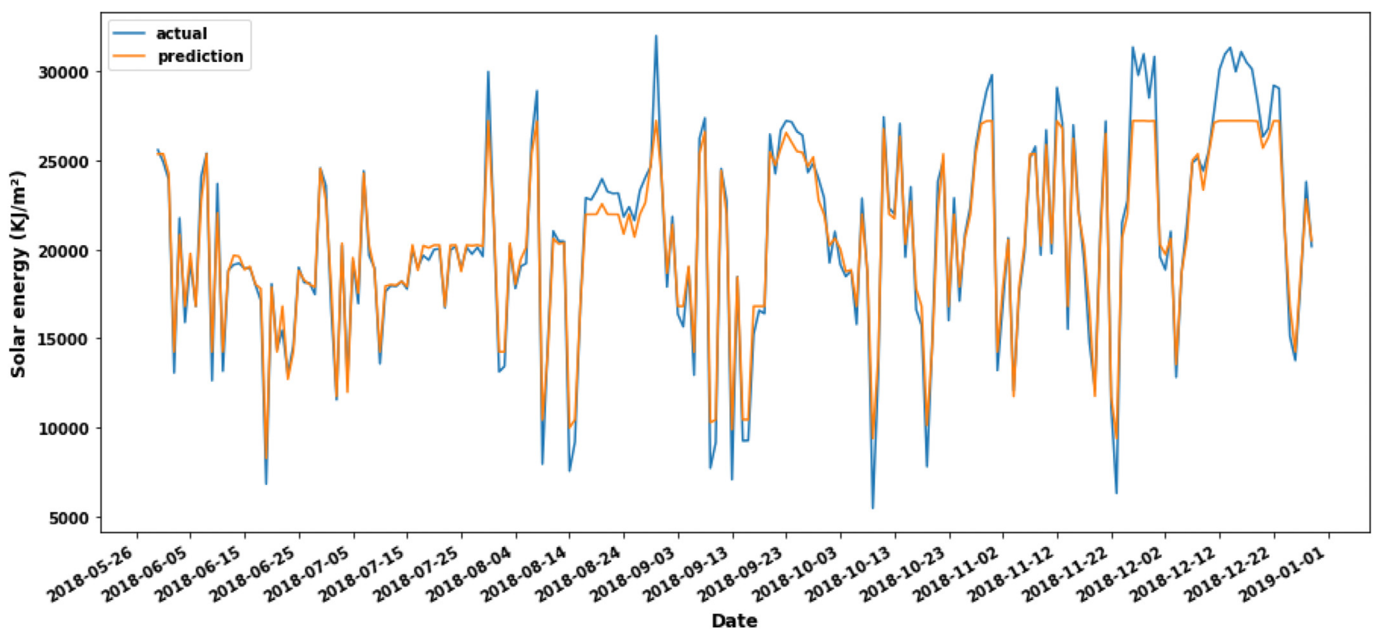

Fig. 26. Real and predicted values of daily solar energy using MLP (Brazil, Pirapora).

range (the maximum value minus the minimum value) of the observed values. Table 5 presents a results synthesis based on this metric for both real-time and daily solar energy prediction and shows the corresponding training time. We can see that RF and ANN report good NRMSE values with respectively, 0.06 and 0.07 for real-time forecast, and 0.03 for daily prediction. Although the daily prediction results of the all studied models are less good than those corresponding to real-time prediction, we can see that only ANN provides very good and promising results for both real-time and daily predictions. Hence, we think that this model seems to be also adapted to long-term solar energy prediction and should be investigated and recommended for this purpose.

To show the quality and reproducibility of our study, we have applied our four models to the case of Brazil, especially Pirapora, a region with a tropical climate that is different from the semi-desert one, and which hosts the largest solar power plant in Latin America. Tables 6 and 7 presents respectively, the hourly and daily predictions of solar energy in the Pirapora region in Brazil, using hourly data available on the Brazilian Institute of Meteorology (INMET) web site [66]. Figs. 19–26 show the curves corresponding to real and predicted values of solar energy using respectively, LR, RF, SVR, and ANN for both hourly and daily predictions. The obtained results compared with those related to the Moroccan's (Errachidia) case (corresponding to the semi-desert climate area), confirm the reproducibility of our study in another country, and more precisely, in another climate area (tropical one). Both RF and ANN perform better with very high accuracy, and ANN still provides good accuracies for both real-time and short-term predictions.

5. Conclusion and perspectives

Solar energy is gradually integrated into modern grids thanks to the availability of more low-cost panels nowadays. Given the expansion of photovoltaic energy, several tools and approaches were proposed in the literature to exploit the potential of AI techniques for solar energy forecasting. However, all of them propose different techniques for separately performing short-term or long-term predictions, and don't focus on real-time forecasting which is crucial to ensure better management and security in the solar energy field. Hence, a unified tool still is interesting to allow performing real-time and short-term as well as long-term predictions, and thus, ensure better management and security as well as maintenance while avoiding the use of several, tools and then the underlying complexity and error issues.

For this purpose, this paper studies the prediction accuracy of four different ML models: LR, RF, SVR, and ANN (MLP) which is a deep learning one since we implement it using 8 learning layers. The models were firstly experimented with data from a Moroccan region, to perform both real-time and short-time solar energy forecasts. The results discussed according to six metrics MAE, MSE, RMSE, Max Error, R-squared, and NRMSE which evaluate both the prediction accuracy and the error range, show (i) the efficiency of using Pearson correlation coefficient to identify the most correlate and relevant meteorological data from which the models should learn, and their great impact on the prediction accuracy and good models' fitting, (ii) the goodness of the accuracy provided by RF and ANN for real-time solar energy prediction thanks to their ability to solve complex and non-linear issues, against LR and SVR which show poor results, (iii) the goodness of the accuracy provided by ANN which outperforms all of the other models for daily solar energy predictions thanks to its ability to perform deep computing using several learning layers, but also due to the context of data which has been aggregated. It was indeed the only model that has provided good results and metrics values. According to these observations, we could conclude that ANN which has reported good accuracies and low errors, is a simple and robust model that is able to minimize errors during the learning process, and is suitable for real-time and short-term solar energy predictions technique which seems to be also promising for long-term solar energy forecasting. Shortly, we think that this work has contributed to identifying an integral solution that should be adopted to implement a unified tool based on the Person correlation coefficient and the deep learning model ANN for both real-time and short-term solar energy forecasting.

As perspectives of this work, we plan to apply our key findings to other different climate areas and to improve our model using other techniques, such as the Resnets network. We note that the experiments of this network in the literature, have shown that the residual structure can efficiently increase the speed of convergence of model training, and can also effectively reduce the problem of degradation. Besides, we also plan to experiment with our key findings for other solar energy forecasting horizons, namely, the medium-term and long-term ones. Note that other promising deep learning models should also be investigated in the perspective of providing a globally optimized prediction accuracy while considering all of the different solar energy forecasting horizons.

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.

References

[1] Sen S, Ganguly S. Opportunities , barriers and issues with renewable energy development — a discussion. Renew Sustain Energy Rev 2016. https://doi.org/10.1016/j.rser.2016.09.137. 0—1.
[2] De Giorgi MG, Congedo PM, Malvoni M. Photovoltaic power forecasting using statistical methods: impact of weather data. IET Sci Meas Technol 2014;8: 90—7. https://doi.org/10.1049/iet-smt.2013.0135.
[3] Wan C, Zhao J, Song Y, Xu Z, Lin J, Hu Z. Photovoltaic and solar power forecasting for smart grid energy management. CSEE J Power Energy Syst 2016;1: 38—46. https://doi.org/10.17775/cseejpes.2015.00046.
[4] Ts IEC. IEC TS 62446-3 technical specification. 2017.
[5] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: a review. Renew Sustain Energy Rev 2018;81:912—28. https://doi.org/10.1016/j.rser.2017.08.017.
[6] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de-Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. Sol Energy 2016;136:78—111. https://doi.org/10.1016/j.solener.2016.06.069.
[7] Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, et al. Machine learning methods for solar radiation forecasting: a review. Renew Energy 2017;105:569—82. https://doi.org/10.1016/j.renene.2016.12.095.
[8] Yadav HK, Pal Y, Tripathi MM. Photovoltaic power forecasting methods in smart power grid. In: 12th IEEE int conf electron energy, environ commun comput control (E3-C3). INDICON; 2015 2016. https://doi.org/10.1109/INDICON.2015.7443522.
[9] Raza MQ, Nadarajah M, Ekanayake C. On recent advances in PV output power forecast. Sol Energy 2016;136:125—44. https://doi.org/10.1016/j.solener.2016.06.073.
[10] Andr M, Calif R, Soubdhan T. Hourly forecasting of global solar radiation based on multiscale decomposition methods. A hybrid approach 2017;119:288—98. https://doi.org/10.1016/j.energy.2016.11.061.
[11] Benali L, Notton G, Fouilloy A, Voyant C, Dizene R. Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renew Energy 2018. https://doi.org/10.1016/j.renene.2018.08.044.
[12] Liu D, Sun K. Random forest solar power forecast based on classification optimization. Energy 2019;187:115940. https://doi.org/10.1016/j.energy.2019.115940.
[13] Wolff B, Kramer O, Heinemann D. Selection of numerical weather forecast features for PV power predictions with random forests. Lect Notes Comput Sci 2017;10097. https://doi.org/10.1007/978-3-319-50947-1_8. LNAI:78—91.
[14] Lahouar A, Mejri A, Ben Hadj Slama J. Importance based selection method for day-ahead photovoltaic power forecast using random forests. Int Conf Green Energy Convers Syst GECS 2017. https://doi.org/10.1109/GECS.2017.8066171.
[15] Tato JH, Brito MC. Using smart persistence and random forests to predict photovoltaic energy production. Energies 2019;12:1—12. https://doi.org/10.3390/en12010100.
[16] Yao W, Zhang C, Hao H, Wang X, Li X. A support vector machine approach to estimate global solar radiation with the in fl uence of fog and haze. Renew Energy 2018;128:155—62. https://doi.org/10.1016/j.renene.2018.05.069.
[17] Fan J, Wang X, Wu L, Zhou H, Zhang F, Yu X, et al. Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates : a case study in China. Energy Convers Manag 2018;164:102—11. https://doi.org/10.1016/j.enconman.2018.02.087.
[18] Ayodele TR, Ogunjuyigbe ASO, Amedu A, Munda JL. Prediction of global solar irradiation using hybridized k-means and support vector regression algorithms. Reinforc Plast 2019;29:78—93. https://doi.org/10.1016/j.ref.2019.03.003.
[19] Palaniappan V. Predicting solar irradiance with SVM regression. SSRN Electron J 2019. https://doi.org/10.2139/ssrn.3302155.
[20] Hu Y, Lian W, Han Y, Dai S, Zhu H. A Seasonal Model Using Optimized Multi-Layer Neural Networks to Forecast Power Output of PV Plants n.d.:1—17. https://doi.org/10.3390/en11020326.
[21] Zhu H, Li X, Sun Q, Nie L, Yao J, Zhao G. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. Energies 2016;9:1—15. https://doi.org/10.3390/en9010011.
[22] Cervone G, Clemente-Harding L, Alessandrini S, Delle Monache L. Short-term photovoltaic power forecasting using artificial neural networks and an analog ensemble. Renew Energy 2017;108:274—86. https://doi.org/10.1016/j.renene.2017.02.052.
[23] Khan I, Zhu H, Yao J, Khan D. Photovoltaic power forecasting based on Elman Neural Network software engineering method. Proc IEEE Int Conf Softw Eng Serv Sci ICSESS 2018;2017. https://doi.org/10.1109/ICSESS.2017.8343021. Novem:747—50.
[24] Alfadda A, Rahman S, Pipattanasomporn M. Solar irradiance forecast using aerosols measurements : a data driven approach. Sol Energy 2018;170: 924—39. https://doi.org/10.1016/j.solener.2018.05.089.
[25] Liu L, Zhan M, Bai Y. A recursive ensemble model for forecasting the power output of photovoltaic systems. Sol Energy 2019;189:291—8. https://doi.org/10.1016/j.solener.2019.07.061.

[26] Paiva G, Pimentel SP. Multiple site intraday solar irradiance forecasting by machine learning algorithms: MGGP and MLP neural networks. Energies 2020;13(11):3005. https://doi.org/10.3390/en13113005.
[27] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98. https://doi.org/10.1016/j.rser.2019.02.006.
[28] Alizamir M, Kim S, Kisi O, Zounemat-kermani M. A comparative study of several machine learning based non-linear regression methods in estimating solar radiation : case studies of the USA and Turkey regions. Energy 2020;197:117239. https://doi.org/10.1016/j.energy.2020.117239.
[29] Deo RC, Ş Mehmet. Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature (LST) for regional locations in Queensland. https://doi.org/10.1016/j.rser.2017.01.114; 2017. 72, 828, 848.
[30] Feng J, Wang W, Li J. An LM-BP neural network approach to estimate monthly-mean daily global solar radiation using. 2018. p. 1–14. https://doi.org/10.3390/en11123510.
[31] Ozoegwu CG. Arti fi cial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number. J Clean Prod 2019;216:1–13. https://doi.org/10.1016/j.jclepro.2019.01.096.
[32] Chaudhary P, Rizwan M. Energy management supporting high penetration of solar photovoltaic generation for smart grid using solar forecasts and pumped hydro storage system. Renew Energy 2017. https://doi.org/10.1016/j.renene.2017.10.113.
[33] Demirdelen T, Ozge Aksu I, Esenboga B, Aygul K, Ekinci F, Bilgili M. A new method for generating short-term power forecasting based on artificial neural networks and optimization methods for solar photovoltaic power plants. Springer Singapore; 2019. https://doi.org/10.1007/978-981-13-6151-7_8.
[34] Obi M, Bass R. Trends and challenges of grid-connected photovoltaic systems – a review. Renew Sustain Energy Rev 2016;58:1082–94. https://doi.org/10.1016/j.rser.2015.12.289.
[35] Li P, Zhou K, Yang S. Photovoltaic power forecasting: models and methods. In: 2nd IEEE conf energy internet energy syst integr EI2 2018 - proc; 2018. p. 1–6. https://doi.org/10.1109/EI2.2018.8582674.
[36] Jiménez-Pérez PF, Mora-López L. Modeling and forecasting hourly global solar radiation using clustering and classification techniques. Sol Energy 2016;135:682–91. https://doi.org/10.1016/j.solener.2016.06.039.
[37] Alfadda A, Rahman S, Pipattanasomporn M. Solar irradiance forecast using aerosols measurements : a data driven approach. Sol Energy 2018;170:924–39. https://doi.org/10.1016/j.solener.2018.05.089.
[38] Ding K, Feng L, Wang X, Qin S. Forecast of PV power generation based on residual correction of Markov chain. 2015. https://doi.org/10.1109/ICCAIS.2015.7338692.
[39] Hossain M, Mekhilef S, Danesh M, Olatomiwa L. Application of extreme learning machine for short term output power forecasting of three grid-connected. PV systems 2017;167:395–405. https://doi.org/10.1016/j.jclepro.2017.08.081.
[40] Carte D'IRRADIATION SOLAIRE du maroc. http://solarforafrica.net/map/. [Accessed 20 March 2020].
[41] Monographie de la province d'Errachidia. 2019. http://www.equipement.gov.ma/Carte-Region/RegionMeknes/Presentation-de-la-region/Monographie/Documents/Monographie%20Errachidia.pdf/. [Accessed 13 November 2019].
[42] Bouaichi A, El Amrani A, Ouhadou M, Lfakir A, Messaoudi C. In-situ performance and degradation of three different photovoltaic module technologies installed in arid climate of Morocco. Energy 2019;190:63–8. https://doi.org/10.1016/j.energy.2019.116368.
[43] Liu L, Liu D, Sun Q, Li H, Wennersten R. Forecasting power output of photovoltaic system using A BP network method. Energy Procedia 2017;142:780–6. https://doi.org/10.1016/j.egypro.2017.12.126.
[44] Wang F, Zhen Z, Mi Z, Sun H, Su S, Yang G. Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. Energy Build 2015;86:427–38. https://doi.org/10.1016/j.enbuild.2014.10.002.
[45] Choi JH. Investigation of the correlation of building energy use intensity estimated by six building performance simulation tools. Energy Build 2017;147:14–26. https://doi.org/10.1016/j.enbuild.2017.04.078.
[46] Zhou H, Deng Z, Xia Y, Fu M. A new sampling method in particle filter based on Pearson correlation coefficient. Neurocomputing 2016;216:208–15. https://doi.org/10.1016/j.neucom.2016.07.036.
[47] Xu H, Deng Y. Dependent evidence combination based on shearman coefficient and Pearson coefficient. IEEE Access 2018;6:11634–40. https://doi.org/10.1109/ACCESS.2017.2783320.
[48] É Biernat, Lutz M. Data science: fondamentaux et études de cas: machine learning avec Python et R. Editions Eyrolles; 2015.
[49] Pavlov YL. Random forests. Random For 2019;1–122. https://doi.org/10.1007/978-3-662-56776-0_10.
[50] Prasad R, Ali M, Kwan P, Khan H. Designing a multi-stage multivariate empirical mode decomposition coupled with ant colony optimization and random forest model to forecast monthly solar radiation. Appl Energy 2019;236:778–92. https://doi.org/10.1016/j.apenergy.2018.12.034.
[51] Abuella M, Chowdhury B. Random forest ensemble of support vector regression models for solar power forecasting. In: 2017 IEEE power energy soc innov smart grid technol conf ISGT. vol. 2017; 2017. p. 2–6. https://doi.org/10.1109/ISGT.2017.8086027.
[52] Raza MQ, Khosravi A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. Renew Sustain Energy Rev 2015;50:1352–72. https://doi.org/10.1016/j.rser.2015.04.065.
[53] Fleetwood A, Galarza A. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. https://doi.org/10.1016/j.renene.2018.03.070; 2018. 126, 855, 864.
[54] Stošović MA, Litovski V. Applications of artificial neural networks in electronics. Electronics 2017;22:87–94. https://doi.org/10.7251/ELS1721087A.
[55] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: a review. Energy Convers Manag 2018;156:459–97. https://doi.org/10.1016/j.enconman.2017.11.019.
[56] Gueymard CA. 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 2014;39:1024–34. https://doi.org/10.1016/j.rser.2014.07.117.
[57] Wang F, Mi Z, Su S, Zhao H. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. Energies 2012;5:1355–70. https://doi.org/10.3390/en5051355.
[58] Squared Miles BJR, Adjusted R. Squared Wiley 2014.
[59] Jebli I, Belouadha FZ, Kabbaj MI. The forecasting of solar energy based on Machine Learning. In: 2020 international conference on electrical and information technologies (ICEIT). IEEE; 2020. p. 1–8. https://doi.org/10.1109/ICEIT48248.2020.9113168.
[60] Lyu L, Kantardzic M, Arabmakki E. Solar irradiance forecasting by using wavelet based denoising. In: Ieee SSCI 2014 - 2014 IEEE symp ser comput intell - CIES 2014 2014 IEEE symp comput intell eng solut proc; 2014. https://doi.org/10.1109/CIES.2014.7011839. 110–6.
[61] Grimaccia F, Leva S, Mussetta M, Ogliari E. Analysis and validation of ANN PV power output forecasting at 24 hours ahead. ElectrIMACS 2014;131:88–100.
[62] Choi H, Ryu S, Kim H. Short-term load forecasting based on ResNet and LSTM. 2018 IEEE Int Conf Commun Control Comput Technol Smart Grids 2018:1–6.
[63] Onyutha C. From R-squared to coefficient of model accuracy for assessing " goodness-of-fits. 2020.
[64] Figueiredo Filho DB, Júnior JAS, Rocha EC. What is R^2 all about? Leviathan 2011;3:60–8.
[65] Mohammed AA, Aung Z. Ensemble learning approach for probabilistic forecasting of solar power generation. Energies 2016;9:10–7. https://doi.org/10.3390/en9121017.
[66] National Institute of Meteorology (inmet). http://www.inmet.gov.br/portal/; 2019.