Ecological Informatics 69 (2022) 101643

Contents lists available at ScienceDirect

Ecological Informatics

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

# Prediction of daily global solar radiation and air temperature using six machine learning algorithms; a case of 27 European countries

Modeste Kameni Nematchoua [a,*], José A. Orosa [b], Marwa Afaifia [c]

[a] Mathematics Department, Universite Libre de Bruxelles (ULB), Belgium
[b] Department of N.S. and M.E. ETSNyM, University of A Coruña, Paseo de Ronda 51, 15011A Coruña, Spain
[c] Ecole Polytechnique d'Architecture et d'Urbanisme (EPAU), Laboratoire Ville, Urbanisme et Développement Durable (VUDD), Route de Beaulieu, El-Harrach - BP N°177, 16200, Algiers, Algeria

ARTICLE INFO

*Keywords:*
Solar radiation
Europe, machine-learning
Prediction

ABSTRACT

The prediction of global solar radiation in a region is of great importance as it provides investors and politicians with more detailed knowledge about the solar resource of that region, which can be very beneficial for large-scale solar energy development. In this sense, the main objective of this study is to predict the daily global solar radiation data of 27 cities (Brussels, Paris, Lisbon, Madrid…), located in 27 countries, which have mostly different solar radiation distributions in Europe. In this research, six different machine-learning algorithms (Linear model (LM), Decision Tree (DT), Support Vector Machine (SVM), Deep Learning (DL), Random Forest (RF) and Gradient Boosted Trees (GBT)) are used. In the training of these algorithms, daily air temperature(Ta), wind speed(Va), relative humidity(RH) and solar radiation of these cities are used. The data is supplied from the Meteonorm tool and cover the last years grouped in two periods (1960–1990; 2000–2019). To decide on the success of these algorithms, four different statistical metrics (Average Relative Error (ARE), Average absolute Error (AAE), Root Mean Squared Error (RMSE), and $R^2$ (R-Squared)) are discussed in the study. In addition, the forecasting of air temperature and global solar radiation of these cities in 2050 and 2100 were made using three of the most recent Intergovernmental Panel on Climate Change (IPCC) scenarios (RCP2.6; RCP 4.5, and RCP 8.5). The results show that ARE, $R,^2$ and RMSE values of all algorithms are ranging from 0.114 to 6.321, from 0.382 to 0.985, from 0.145 to 2.126 $MJ/m^2$, respectively. By analysing all the algorithms, it is noticed that the Decision tree exhibited the worst result in terms of $R,^2$ and RMSE metrics. Among the six prediction algorithms, the DL was recognized as the only algorithm that exceeded the t-critical value (The t-critical value is the cutoff between retaining or rejecting the null hypothesis). Globally, all the six machine learning algorithms used in this research can be applied to predict the daily global solar radiation data with good accuracy. Despite this, the SVM model is the best model among all the six models used. It is followed by the DL, LM, GB, RF and DT, respectively.

## 1. Introduction

The Sun, like the wind, the sea, and biomass, are free and sustainable energy resources. The impacts of the sun can be positive (favourable to fauna, flora and human health), or negative, because a high density of the sun can cause famine, drought, and certain diseases harmful to human health. Solar radiation is considered one of the most important sources of energy on Earth. Indeed, it plays a very significant role in the surface radiative balance, meteorological and climatic extremes, and the photosynthesis of vegetation (Huang et al., 2001).

Solar radiation is the set of electromagnetic waves emitted by the Sun (Sullivan et al., 2017). It is made up of the full range of radiation, from far-ultraviolet such as gamma rays to radio waves via visible light (Sullivan et al., 2017). The solar radiation received on the ground varies over time, on the one hand as a function of variations in solar activity, and the other hand as a function of the seasons (according to the inclination of the Earth) and within each season according to the natural and anthropogenic variations in cloudiness (CNRS, 2020). To evaluate the solar deposit of a city, it is necessary to have data relating to the solar radiation and other climatic parameters of this city (Yadav et al., 2014; Yıldırım et al., 2018). For good planning of the investments in the production of solar energy, it is important to evaluate and predict the amount of solar radiation arriving over the entire energy production area (Balog et al., 2019).

* Corresponding author.
*E-mail address:* kameni.modeste@yahoo.fr (M.K. Nematchoua).

https://doi.org/10.1016/j.ecoinf.2022.101643
Received 23 January 2022; Received in revised form 19 March 2022; Accepted 6 April 2022
Available online 12 April 2022

It is very important to make careful forecasting of solar radiation, long and short term if the energy is produced following an investment intended for marketing. Making precise predictions has a very significant impact on energy beneficiaries and suppliers (Gürel et al., 2020; Jiang et al., 2017). In addition, the accurate prediction of daily global solar radiation is also very relevant both in climate research and in the solar industry. According to Beer et al. (Beer et al., 2010), solar radiation has an important role in the global climate. Indeed, a few small changes in the energy production of the Sun cause considerable changes in the climate of the Earth. According to Garland et al. (Garland et al., 1990), analyses and observations of the variability of solar radiation are important in the research of building materials and extreme weather events.

Solar radiation is studied using sophisticated and expensive instruments, regularly demanding maintenance and calibration. Therefore, very few weather stations in the world measure this climatic parameter. In general, measuring global solar radiation everywhere is often not possible, as it requires expensive, time-consuming and precise processes. In addition, radiation values cannot be accurately measured in most countries because measurements can only be made in certain areas. For this reason, it is easier to use meteorological data to calculate the different values of global radiation (Garland et al., 1990). Some parameters (air temperature, relative humidity, wind speed etc.), have a significant effect on the global solar radiation in a region. Nowadays, accounting for over 10,221 meteorological stations in Europe and neighboring countries, very few of them can measure solar radiation (Meteorological data from ground stations, n.d.). Examples, for a total of 1798 meteorological stations identified in Turkey in 2020, only 129 of them are known to measure the values of solar radiation (TSMS, 2020). On 112 meteorological stations identified in Belgium, only 12 measure solar radiation data (Station météo, n.d.; Meteo Belgium, n.d.). In addition, in another country as in China, until 2012, 122 stations were capable to save the solar radiation data on a total of 756 inventories (Zang et al., 2012). Given these data, it is important to predict solar radiation data under basic of other climatic data regularly measured in most weather stations such as; wind speed, relative humidity, air temperature etc.

In the literature, several models are proposed to predict solar radiation. The empirical models are derived from mathematical formulas (Fan et al., 2018a). Indeed, they are easy to use. However, these models are unable to predict with good precision the various data of solar radiation in the short term because of the regular changes of the meteorological conditions (Jahani and Mohammadi, 2019). Nowadays, with the evolution of technology, artificial intelligence (AI) has gained momentum. It is applicable in most fields of engineering (Long et al., 2014). In addition to empirical models, various artificial intelligence models such as Linear model (LM), Decision Tree (DT), Support Vector Machine (SVM), Deep Learning (DL), Random Forest (RF), and Gradient Boosted Trees etc. are models regularly applied for forecasting solar radiation data. Several results from many previous studies show that AI algorithms give more reliable results than those of empirical models for predicting solar radiation (Bayrakçı et al., 2018; Liu et al., 2020).

For example, in 2021, Huang et al. (Huang et al., 2021) used 12 machine learning models (random forest, Gaussian process regression (GPR), gradient boosting regression tree (GBRT) etc.), to predict daily and values of solar radiation. The results showed that climatic parameters such as visibility, sunshine duration, and surface temperature are very significant in the machine learning algorithms. In their research, Chen et al. (Chen et al., 2011) explained that AI algorithms showed more accurate results than those of the empirical models for the prediction of solar radiation. Sun et al.(Sun et al., 2016) in 2016; Persson et al. (Persson et al., 2017) in 2017, Fan et al.(Fan et al., 2018b) in 2018; Zeng et al.(Zeng et al., 2020) in 2020; used tree machine learning algorithms such as the gradient boosting regression tree and the random forest algorithm to predict solar radiation in many regions. The results showed a strong dependence of solar radiation from the meteorological data. In 2020, Abdurrahman et al. (Guher et al., 2020) evaluated the hourly average solar radiation of two geographic locations on the same latitude by using several Machine Learning (ML) algorithms such as K-Nearest Neighbors ($K-$NN), Support Vector Machines, Multilayer Feed-Forward Neural Network (MFFNN) etc. The input variables were applied by using 6 different feature selection methods with Waikato Environment for Knowledge Analysis (WEKA) software. The MFFNN models were found as the most successful estimation models regarding the RMSE, MAE (mean absolute error), and R,$^2$ which were 0.0508–0.0536, 0.0341–0.0352, and 0.9488–0.9656, respectively. In 2017, Quej et al. (Quej et al., 2017a) used several machine-learning algorithms such as SVM, ANN (Artificial Neural Network) for predicting daily global solar radiation data in six regions in Mexico, with input data, extra-terrestrial solar radiation, air temperature, rainfall. The best results were obtained in SVM with RMSE $= 2.578$, MAE $= 1.97$ and $R^2 = 0.689$. In another research, Meenal and Selvakumar (Meenal and Selvakumar, 2018), using the SVM, and ANN models for predicting the daily global solar radiation. The best accuracy was obtained by applying SVM (with $R^2$ around 0.99).On the other hand, Marzo et al. (Marzo et al., 2017) predicted the daily global solar radiation of 13 regions using minimum temperature, and maximum temperature, extra-terrestrial solar radiation for training the ANN algorithm. The best prediction result was obtained whereas $R^2$ was estimated to 0.64, and Relative *Mean* Bias Errors (RMSE) $<$ 4%. Neelamegam and Amirtham (Neelamegam and Amirtham, 2016) used two ANN models with four different algorithms for predicting the daily global solar radiation of five different locations across India, with the meteorological data collected for the last 10 years. The results of this research confirmed that the prediction accuracy of the ANN model depends on the total dataset used to train the network for the new application.

Several methods are regularly used in the prediction of global solar radiation. Most of them give results that are close to each other for the study sites having the same climatic variations. However, the different results obtained and the data accuracy of prediction vary according to the input parameters used and the volume of data used. The purpose of this research is to predict daily global solar radiation data for 27 cities, of which the majority have a different solar radiation distribution in the Europe. In addition, this research assesses the evolution of air temperature in 2050 and 2100 using three IPCC scenarios (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.).

We deduce from the previous studies that the applied algorithms generally give close results to each other for the different sites analysed. Therefore, the implementation of some metrics to discuss the performance of the algorithms may be incomplete. With these different points of view, this study contributes to the literature in the following three ways:

- Analysis and comparison of six frequently and rarely used machine learning algorithms on the same dataset: in addition to the frequently used Linear model, Decision Tree, and Support Vector Machine algorithms, Deep Learning, Random Forest and Gradient Boosted Trees are also used in the prediction of daily global solar radiation data in this study.

- Prediction of the global solar radiation distribution of 27 cities located in 27 countries in Europe, with very low, low, medium and high solar radiation potentials. These cities represent the global solar radiation distribution in Europe.

- A detailed discussion of all the results: Indeed, unlike other research that focuses on the comparison of different prediction successes with several metrics, this research presents a complete discussion of the success of the algorithm using four metrics together (Average Relative Error, Average absolute Error, Root Mean Squared Error, and R-Squared).

The content of this work is spread over four sections. Section 1 presents a review of the literature on the various studies linked to the prediction of solar radiation; Section 2 explains in detail the main methodology used in this research; Section 3 presents the results and

**Table 1**
Geo-localization of some cities in Europe.

| Country | Capital | Latitude | Longitude |
|---|---|---|---|
| Germany | Berlin | 52°31 N | 13°24E |
| Belgium | Brussels | 50°51 N | 4°20E |
| Bulgaria | Sofia | 42°41 N | 23°19E |
| Croatia | Zagreb | 45°48 N | 15°58E |
| Danemark | Copenhague | 55°40 N | 12°33E |
| Spain | Madrid | 40°24 N | 3°42O |
| Finland | Helsinki | 60°10 N | 24°56E |
| France | Paris | 48°51 N | 2°20E |
| Greece | Athens | 37°58 N | 23°42E |

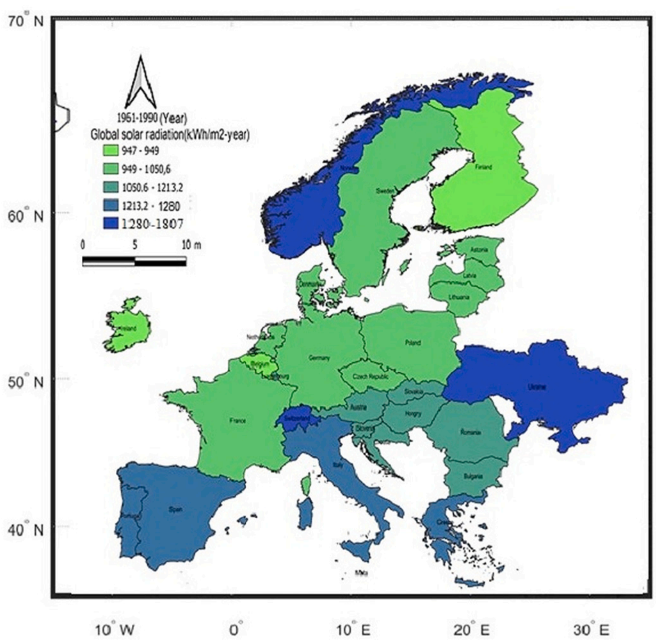


**Fig. 1.** Daily global solar radiation of European countries from 1961 to 1990.

analyses the daily global solar radiation in the 27 countries studied; and finally, section 4 shows some comparisons and discussions of the results obtained.

## 2. Methodology

### 2.1. Study sites

The European covers an area over 4.2 million square kilometers, is populated by more than 446 million inhabitants and is the third-largest economic power in the world. The global solar radiation potential on the European countries is used in spite of some variations in the sunlight concentration between each city (CNRS, 2020). Because of the geographical location, the solar energy potential of Southern Europe is higher than in Northern Europe. The annually sunshine duration is estimated around of 2130 h, whereas, the average solar radiation is estimated to be between 1080 kWh/m$^2$ and 2200 kWh/m$^2$.

In this research, the 27 capitals of some countries in Europe were analysed. Main details and the view of some countries in the European map are given in Table 1 and also in Fig. 1, respectively.

In Table 1, we have selected at random some countries of the Europe that we present their geographic coordinates. We see through Fig. 1 that the daily global solar radiation of the last 30 years (from 1961 to 1990), of the 27 countries of the European can be represented on different scales. It is noticed that solar radiation is a little lower in Belgium, Ireland, and Finland. In addition, Portugal, Italy, Greece, and Spain can

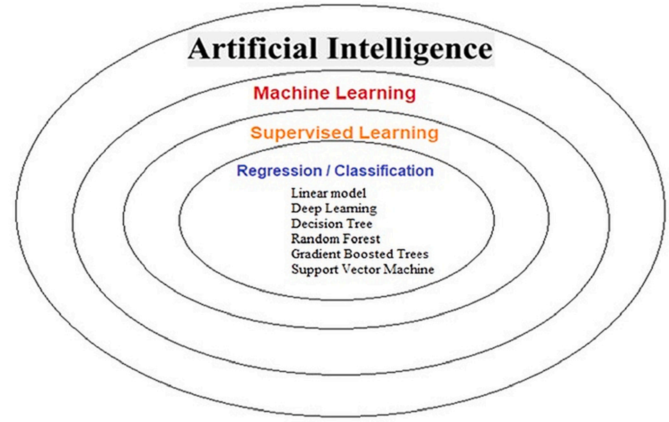


**Fig. 2.** The core structure of the study.

be considered transition countries in terms of solar radiation. Finally, Norway and Ukraine are countries with the highest daily global solar radiation. Based on these results, we infer that the study conducted in these 27 countries can be easily adopted in most of the cold, slightly cold, and warm climate countries of the world. Finally, it must be clarified that, in this study, the term "solar radiation" will always refer to "global solar radiation at the horizontal surface".

### 2.2. Data collection

The data processing includes mapping of different global solar radiation and air temperature scenarios for 2050 and 2100. These scenarios are developed for comparison at the level of the European countries.

The daily data of air temperature, global solar radiation, wind speed and relative humidity from the last years (2000–2019) were applied. These data come from 27 meteorological stations located in the 27 countries studied, and measured at a height of 100 m. These data can be easily downloaded with Meteonorm software. In addition to this, the daily global solar radiation applied in this research were calculated under basic of eqs. (1)–(4) (Bakirci, 2009), following:

$$\mathrm{H\_Gh} = \frac{24}{\pi} I_{sc}\left(1 + 0.033cos\frac{360D}{365}\right) * \left(cos(\varphi)cos(\delta)sin(ws)\frac{2\pi ws}{360}sin(\varphi)sin(\delta)\right) \quad (1)$$

Where $I_{sc}$ is a constant of solar fixed at 1367 W/m$^2$, while the variable φ is set as the latitude angle, in addition, δ is considerate as the declination angle, $\omega s$ is the angle of the sunset times, and $D$ is the number of days beginning on 1st January. The eqs. (2,3 and 4) allow to calculate the sunset hour angle and declination angle(Bakirci, 2009).

$$\delta = 23.45sin\left[\frac{360(D + 284)}{365}\right] \quad (2)$$

$$ws = cos^{-1}[-tan(\delta)tan(\varphi)] \quad (3)$$

The day length has been calculated according to the formula given in (Huang et al., 2001):

$$S_o = \frac{2}{15}ws \quad (4)$$

### 2.3. Machine learning algorithms

Machine learning (ML) gives the systems with the possibility to understand by itself and then to evaluate some outputs (Guher et al., 2020; Huang et al., 2021). The performance of a machine learning algorithm is related to its selection of attributes and also to the success of the learning process. In this research, six ML algorithms were applied. As detailed in

**Table 2**
Summary of the statistical metrics.

| Metrics | Equation | Description and reference |
|---|---|---|
| Average absolute error (AAE) | $\frac{\sum_{i=1}^{n}\lvert y_i - x_i \rvert}{n}$ | AAE performance is important of prediction metric models for the long-term small. A value nearby 0 is an indication of better performance (Fan et al., 2018c) |
| Root mean squared error (RMSE) | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$ | RMSE gives news regarding the short-term accuracy of prediction models. RMSE should be positive and nearby 0 for a good accuracy (Fan et al., 2018c) |
| Coefficient of determination ($R^2$) | $1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \overline{x_i})^2}$ | Value from 0 to 1. A good performance main that $R^2$ should be nearby 1 (Rehman, 1998). |

δ = percent error, υA = actual value observed, υE = expected value

Fig. 2, these are linear model (LM), Decision Tree (DT), Support Vector Machine (SVM), and Deep Learning (DL), Random Forest (RF), and Gradient Boosted Trees. It is interesting to note that the different algorithms were executed in R-Studio, software version 4.0.5. The input data consisted of daily air temperature, wind speed, relative humidity and global solar radiation. The data set was divided in two: training (70%) and testing (30%). The different types of errors detailed in Table 2 give an overview of the precision between measured and simulated data.

### 2.3.1. *Linear model*

Linear regression (LR) is considered one of the most known and well-understood algorithms in machine learning (Machine Learning, 2019). Linear regression is recognized as a linear approach to perform the strong link between one or more explanatory variables and a scalar response (Freedman, 2009a; Freedman, 2009b; Rencher and Christensen, 2012; Seal, 1967). Some good information regarding this model can be found in (Freedman, 2009a). In this research, the best results are observed in which ($R^2$ around 0.9) and (RMSE<10) parameters, for example in Roma, Sofia, Warsaw and Stockholm.

### 2.3.2. *Decision tree*

DT algorithms are commonly used in ML, in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a purpose, Such as the Decision Tree algorithm, It is also a popular tool in machine learning (Kamiński et al., 2017; Karimi and Hamilton, 2011; Quinlan, 1987a; Quinlan, 1987b). Some information regarding this model can be found in the literature, and more precise in the references (Quinlan, 1983; Wagner, 1975).

### 2.3.3. *Random Forest*

Random Forests (RF) is an ensemble learning technique proposed by Breiman which combines bagging and random feature selection by using a large number of non-pruned decision trees. Hence, each individual tree is trained on a different subset of samples (due to bagging), as well as a different subset of features (due to random feature selection) the random feature selection for every tree allows to decorrelate the predictions of the different trees (Ho, 1995). The aggregation of decorrelated classifiers allows reducing the variance of the final prediction. In the case of classification, the aggregation is performed by majority vote (i.e. the class that is predicted by the largest number of individual classifiers is selected as the prediction) (Ho, n.d.). In the case of regression, the aggregation is performed by performing the average of the individual predictions (Maillo et al., n.d.; Saikia et al., n.d.; Erdal and Aytug, 2016; Nordhaug et al., 2018).

### 2.3.4. *Support vector machine (SVM)*

The machine learning algorithm named support vector machine is commonly used to handle classification and regression problems (Min et al., 2003). In such an algorithm, each data item constitutes a specific point located in an n-dimensional space (n in this case represents the number of features obtained). The content of each feature represents the value of a particular point.

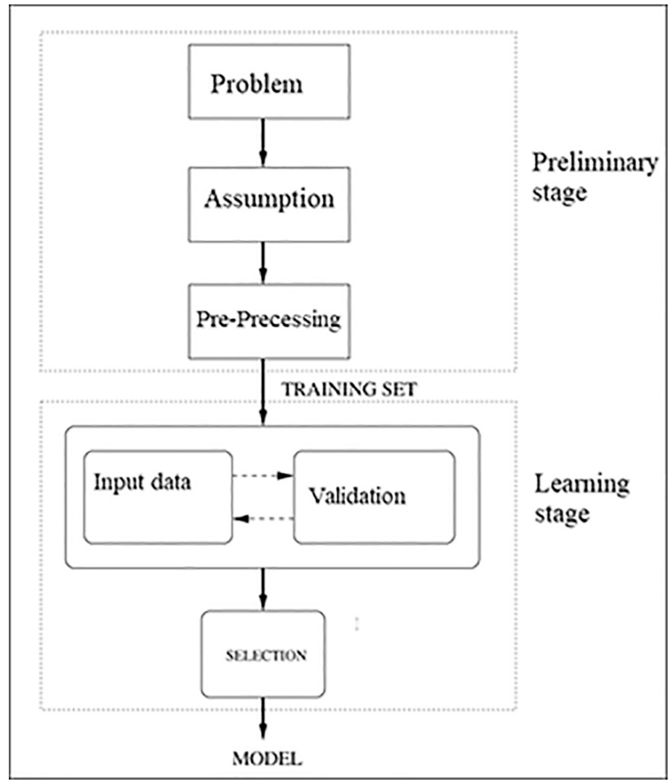


**Fig. 3.** Preliminary and learning stage of data.

### 2.3.5. *Deep learning (DL)*

DL is considered as one of the very popular programming tools applied to search for a solution to somewhat complex problems that contain very large data sets (Najafabadi et al., 2015; Pardis et al., 2019; Talib et al., 2016). Thus, in a DL method, so-called semi-supervised, supervised, and unsupervised algorithms are used (Najafabadi et al., 2015). One of the advantages of DL compared to other classical machine learning approaches is that it has the ability to perform feature extraction itself, even if some raw data is known as input and, in this particular approach, an increase in the dataset directly implies an increase in the various learning performances (Morgan. Nelson., 2011; Shaohui et al., 2017; Yi-zhou et al., 2017; Yuming et al., 2015).

### 2.3.6. *Gradient boosted trees*

Boosting is an additional generic ensemble technique that attempts to boost the accuracy of any given learning algorithm (Piryonesi and El-Diraby, 2020). The focus of boosting methods is to produce a series of weak learners in order to produce a powerful combination (Hastie et al., 2009; Piryonesi and El-Diraby, 2021). A weak learner is a learner that has accuracy only slightly better than chance. Unlike Bagging (employed in Random Forests), the resampling of the training set is dependent on the performance of the earlier classifiers, which prevents a parallel implementation of the assembling procedure (Meteonorm, n.d.). In this research, the best predictions of solar radiation with the Gradient Boosted algorithm was observed in Sofia, Prague, and Amsterdam. The different stages are showed in the Fig. 3.

The selection of these six ML algorithms was not done random, indeed, These algorithms are used frequently in many researches to predict daily global solar radiation in a region (Bayrakçı et al., 2018; Huang et al., 2001). These algorithms can be used for numerous

**Table 3**
Chi-Square Tests for Global Solar Radiation (compare to Ta, Va, and RH).

| | Chi-Square Tests | Value | df | Asymp. Sig.(2-sided) |
|---|---|---|---|---|
| Relative Humidity | Pearson Chi-Square | 354374.145[a] | 249243 | 0.061 |
| | Likelihood Ratio | 31308.844 | 249243 | 1.000 |
| | Linear-by-Linear Association | 1595.274 | 1 | 0.000 |

functions such as prediction, curve fitting and regression. The advantage of most of this algorithms is that they do not, always need the knowledge of mathematical calculations between the parameters but they involve lesser computational effort and provide a compact solution for multi-variable issues.

### 2.4. Evaluation metrics

In this study, some metrics are applied such as AAE (Average absolute error), RMSE (Root mean squared error), ARE (Average relative error), and $R^2$ (Coefficient of determination), to compare the performance success of the prediction models. These different statistical metrics, their corresponding equations, and brief descriptions are given in Table 2.

### 2.5. Data pre-processing

Following are six different steps involved in machine learning to perform data pre-processing applied in this research:

**Step 1:** Import Libraries, A library is a collection of modules that can be called and used. In R, we have a lot of libraries that are helpful in data pre-processing (e.g. Ggvis, Plotly, Rcharts, Rbokeh, Broom,StringR, Magrittr, Slidify,Rvest, Future, RMySQL, RSQLite);

**Step 2:** Import data, our next step is to load the collected data. Mostly the datasets are available in CSV formats as they are low in size which makes it fast for processing. So, to load a csv file using the read_csv function of the R's library. Once the dataset is loaded, we have to inspect it and look for any noise. To do so we have to create a feature matrix X and an observation vector Y with respect to X;

**Step 3:** Checking for missing values, Once we created the feature matrix we are looking the missing values.

**Step 4:** Checking for categorical data, Data in the dataset has to be in a numerical form so as to perform computation on it.

**Step 5:** Feature Scaling, Feature scaling is a technique that is used to bring the data value in a shorter range.

**Step 6:** Splitting data into training, validation and evaluation sets: Finally, we need to split our data into different sets, training set to train the model, validation set to validate the accuracy of our model and finally test set to test the performance of our model on generic data. Before splitting the Dataset, it is important to shuffle the Dataset to avoid any biases.

### 2.6. Chi-Square tests

This subsection determines the relevant inputs which more correlate with global solar radiation. As in the majority of the most common applied used for the *t*-test, in this part, all the analysis were carried out with 95% of confidence level (CL) which was considered level of significance equal to 5%. The different analysis was carried out with Chi-Square method. This test is very important in this statistical study because it makes it easier to evaluate or compare two groups (or two measures) and easily take the best possible decision. With the SPSS software, the interval of significance can be freely selected; indeed, it is a quantitative estimate (called the *p*-value) of the probability that the observed differences are random. Results are showed in Table 3. All the statistical analyses were carried out by means of IBM SPSS 24.0 Statistical software. Table 3 shows the Chi-Square test results for global solar radiation with regard to the air temperature, relative humidity and air speed analysis. It can be inferred from Table 3 that there was no obvious difference between the global solar radiation and (Ta, Va, and RH) ($P > 0.05$), as highlighted, the test is significant for $p < 0.05$ (5%); in this case, being the level of significance (*p*-value) equal to 0.061 ($0.061 > 0.05$).

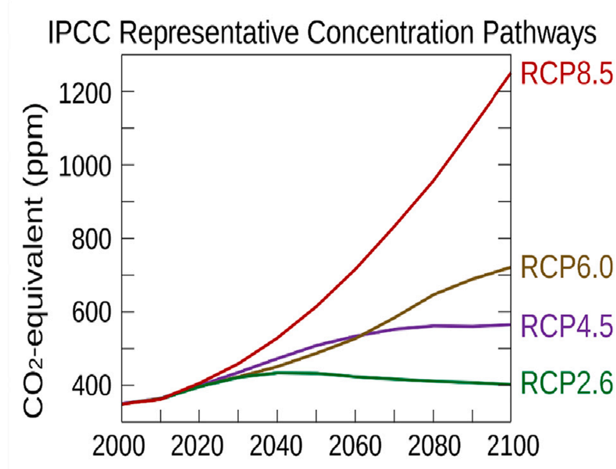


**Fig. 4.** Variation of $CO_2$ as main consequence of the increase of the air temperature (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.).

### 2.7. Scenarios

In this research, it was used three scenarios proposed by the IPCC (RCP 2.6, RCP 4.5, and RCP 8.5), to assess the evolution of the outside temperature and of the global solar radiation in the 27 countries of the European Union. Based on the hourly data for the last thirty years (from 1961 to 1990), we have predicted the evolution of these 2 parameters in 2050 and 2100. The date from (2000–2019) were applied for making prediction with machine Learning algorithm. In each country, only the capitals were selected, because according to the literature, in most countries the capitals are places with a high concentration of population and high pollution (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.; Pielke and Roger, 2021).Fig. 4. shows the variation of carbon according to four scenarios.

#### 2.7.1. Scenario-RCP 2.6

The RCP 2.6 scenario is a "very stringent" pathway. According to the IPCC, the RCP 2.6 scenario implies that CO2 emissions decrease from 2020 onwards and become almost zero in 2100 (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.; Pielke and Roger, 2021). This scenario also requires that methane emissions decrease by half compared to the amount emitted in 2020, and that sulphur dioxide (SO2) emissions decrease by 10% compared to the period 1980–1990 (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.; Pielke and Roger, 2021). In addition, RCP 2.6 requires negative CO2 emissions (e.g. CO2 absorption by trees). The RCP 2.6 scenario conditions negative emissions to 2 gigatons of CO2 per year. In this perspective, the main goal of the RCP2.6 scenario is to keep the global temperature increase below 2 °C by 2100 (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.; Pielke and Roger, 2021).

**Table 4**
AR5 global warming increase (°C) projections (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.).

| SSP Scenario | Range of Global Mean Temperature Increase (°C) - 2100 from pre-Industrial baseline |
|---|---|
| RCP 2.6 | 1.5–2.5 |
| RCP 4.5 | 2.5–3.5 |
| RCP 8.5 | 3.5–5.5 |

#### 2.7.2. Scenario-RCP 4.5

The RCP 4.5 scenario is showed by the IPCC as an intermediate scenario. The peak of emissions in RCP 4.5 will be achieve probably 2040, then will decrease. Seen the conclusions of the IPCC, the RCP 4.5 scenario requires that the CO2 concentration start decreasing in 2045 to reach roughly half of the levels of 2050 by 2100(https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.; Pielke and Roger, 2021).

#### 2.7.3. Scenario- RCP 8.5

In the RCP 8.5 scenario, emissions are expected to continue to increase in the 21st century. This scenario is known to be the most unlikely, but may come true (https://en.wikipedia.org/wiki/Representative_Concentration_Pathway, n.d.). The RCP8.5 scenario is known as the benchmark for worst climate change scenarios (Pielke and Roger, 2021).

The variation of air temperature according to the IPCC is given in Table 4.

### 2.8. Software program language

#### 2.8.1. Algorithm

All the algorithms used in this research were programmed in R-studio, and also sometimes with Jupiter notebook (in python). A total of 6 algorithms were written corresponding to the six MLs applied in this study. It is interesting to note that the performance of a Machine Learning (ML) algorithm is analysed based on the accuracies obtained and the Model Building time (MBT). Model building is the process of developing a probabilistic model that best describes the relationship between the dependent and independent variables.

#### 2.8.2. Simulation software

One of the tools used in this research for predicting solar radiation is Meteonorm version 8. We used the most recent version of this software. Meteonorm generates accurate and representative typical years for any place on earth. It can be chosen from more than 30 different weather parameters (Bontempi, 2021). The database consists of more than 8000 weather stations, five geostationary satellites and a globally calibrated aerosol climatology (Bontempi, 2021). On this basis, sophisticated interpolation models, based on more than 30 years of experience, provide results with high accuracy worldwide (Bontempi, 2021). Meteonorm includes two of the best minute models on the market for reliable simulations of large PV plants or energy management & battery systems. It can model urban heat effects to support the development of green cities. It contains algorithms to calculate extreme years, for example, to test design limits (Bontempi, 2021). It can be even simulated Climate Change using IPCC scenarios. Historical hourly values of irradiation, temperature, humidity, wind and precipitation from 2010 to the present, constantly updated (Bontempi, 2021).

The cartography makes it possible to visualise the spatial-temporal variability of the different IPCC's scenarios of global radiation and air temperature in the long term.

#### 2.8.3. Spatial-temporal visualization

All of the 27 countries of the European Union are characterised by the global radiation and air temperature of 2050 and 2100. The scenarios are used for spatial-temporal visualization through mapping. The cartography is executed according to different scenarios aggregations, using the Quantum GIS-QGIS 3.10.8 software (Qgis, 2021). In this study, the division method of the natural threshold classification was used with five classes, with the aim of grouping data with the same similarities, while maximising the differences between the classes. This division method allows reducing the variance within classes and maximising the variance between the classes.

## 3. Results

### 3.1. Prediction with machine learning algorithms

This research evaluates the predictability of daily solar radiation and the air temperature in the 27 capitals of the European Union via six different machine-learning algorithms. To assess the performance of these algorithms, four different statistical metrics, frequently used in the literature, are discussed.

Table 5 shows the different numerical values of the metrics calculated for the majority of cities studied and the algorithms of the study. As shown in Table 5, the coefficient of determination ($R^2$) varies between 0.5 and 1.0 depending on the city and the algorithm, in other hang, one can easily deduce under the basis of results mentioned in Table 5 that the majority of algorithms in terms of $R^2$ show a good performance in the prediction of daily global solar radiation. In this section, the algorithms related to the studied cities will be analysed considering Table 5 as a reference.

In Paris, Linear Model accomplished an average absolute error of 0.130. This means if you predict daily global solar radiation (H_Gh = 0.60) as a value, the real value is likely between 0.47 and 0.73. The $R^2$ statistic is 0.766 what means that 76.6% of the data fit the regression model. Generally, this value is considered a strong effect size. The deep learning algorithm shows an average absolute error of 0.125. This means if you predict (H_Gh = 0.60) as a value, the real value is likely between 0.475 and 0.725. The $R^2$ statistic is 0.731. This result reveals that 73.1% of the data fit the regression model. Generally, this is a good value. In addition, the Decision Tree algorithm gives an average absolute error of 0.123. This means if you predict (H_Gh = 0.60) as a value, the real value is likely between 0.477 and 0.723. The $R^2$ statistic is 0.714. This result reveals that 71.4% of the data fit the regression model. This value is generally considered a strong effect size (Modeste et al., 2020; Nematchoua et al., 2019f).

In Berlin, Linear Model accomplished an average absolute error of 1.569. This means if you predict daily global solar radiation (H_Gh = 4.04) as a value, the real value is likely between 2.471 and 5.609. The $R^2$ statistic is 0.382. This result reveals that 38.2% of the data fit the regression model. This value is generally considered a weak or low effect size. The Deep Learning model accomplished an average absolute error of 1.166. This means if you predict (H_Gh = 4.04) as a value, the real value is likely between 2.874 and 5.205. The $R^2$ statistic is 0.596. This result reveals that 59.6% of the data fit the regression model. This value is generally considered a moderate effect size. The Decision Tree algorithm accomplished an average absolute error of 1.127 what means that if you predict H_Gh = 4.04 as a value, the real value is likely between 2.913 and 5.167. The $R^2$ statistic is 0.635. This result reveals that 63.5% of the data fit the regression model. This value is generally considered a moderate effect size. In addition, the Random Forest model gives an average absolute error of 1.197. This means if you predict H_Gh = 4.04 as a value, the real value is likely between 2.843 and 5.237. The $R^2$ statistic is 0.594. This result reveals that 59.4% of the data fit the regression model. This value is generally considered a moderate effect size. On the other hand, the Gradient Boosted Trees model accomplished an average absolute error of 1.192. This means if you predict H_Gh = 4.04 as a value, the real value is likely between 2.848 and 5.232. The $R^2$ statistic is 0.602. This result reveals that 60.2% of the data fit the regression model. This value is generally considered a moderate effect

**Table 5**
Performance comparison of cities for each algorithm.

| City | Metric | Linear model | Deep Learning | Decision Tree | Random Forest | Gradient Boosted Trees | Support Vector Machine |
|---|---|---|---|---|---|---|---|
| Paris | RMSE (kWh/m$^2$) | 0.166 | 0.155 | 0.162 | 0.160 | 0.163 | 0.145 |
| | AAE | 0.126 | 0.120 | 0.126 | 0.127 | 0.130 | 0.116 |
| | ARE (%) | 54.74 | 45.80 | 52.84 | 57.53 | 58.37 | 44.00 |
| | $R^2$ | 0.681 | 0.723 | 0.694 | 0.703 | 0.693 | 0.775 |
| | MBT (S) | 7.296 | 1.066 | 0.157 | 2.468 | 13.909 | 38.031 |
| Berlin | RMSE (kWh/m$^2$) | 1.882 | 1.480 | 1.405 | 1.480 | 1.470 | 1.488 |
| | AAE | 1.569 | 1.166 | 1.127 | 1.197 | 1.192 | 1.154 |
| | ARE (%) | 13.7 | 74.67 | 78.94 | 83.47 | 81.03 | 67.64 |
| | $R^2$ | 0.382 | 0.596 | 0.635 | 0.594 | 0.602 | 0.601 |
| | MBT(S) | 4.187 | 0.932 | 0.114 | 0.481 | 8.615 | 15.403 |
| Prague | RMSE (kWh/m$^2$) | 0.188 | 0.190 | 0.197 | 0.181 | 0.179 | 0.185 |
| | AAE | 0.147 | 0.152 | 0.157 | 0.146 | 0.144 | 0.140 |
| | ARE (%) | 74.62 | 81.25 | 84.00 | 80.07 | 76.50 | 67.72 |
| | $R^2$ | 0.716 | 0.712 | 0.690 | 0.733 | 0.741 | 0.712 |
| | MBT(S) | 4.382 | 0.728 | 0.114 | 1.836 | 10.109 | 35.996 |
| Athens | RMSE (kWh/m$^2$) | 0.159 | 0.156 | 0.159 | 0.155 | 0.166 | 0.147 |
| | AAE | 0.125 | 0.125 | 0.129 | 0.127 | 0.137 | 0.118 |
| | ARE (%) | 25.43 | 24.15 | 26.42 | 24.89 | 27.74 | 22.42 |
| | $R^2$ | 0.752 | 0.755 | 0.748 | 0.759 | 0.720 | 0.782 |
| | MBT(S) | 7.623 | 0.741 | 0.112 | 0.991 | 10.504 | 31.37 |
| Lisbon | RMSE (kWh/m$^2$) | 2.126 | 1.429 | 1.432 | 1.432 | 1.482 | 1.449 |
| | AAE | 1.883 | 1.150 | 1.151 | 1.155 | 1.218 | 1.159 |
| | ARE (%) | 70.78 | 38.39 | 39.41 | 39.63 | 40.26 | 36.42 |
| | $R^2$ | 0.373 | 0.649 | 0.649 | 0.646 | 0.616 | 0.630 |
| | MBT (S) | 0.11 | 0.708 | 0.086 | 0.889 | 8.275 | 15.235 |
| Madrid | RMSE (kWh/m$^2$) | 2.065 | 1.485 | 1.503 | 1.505 | 1.444 | 1.459 |
| | AAE | 1.823 | 1.220 | 1.217 | 1.217 | 1.199 | 1.147 |
| | ARE (%) | 75.53 | 40.62 | 41.37 | 41.07 | 38.34 | 40.04 |
| | $R^2$ | 0.537 | 0.634 | 0.600 | 0.600 | 0.624 | 0.629 |
| | MBT(S) | 0.109 | 0.663 | 0.076 | 0.397 | 8.53 | 11.808 |
| Brussels | RMSE (kWh/m$^2$) | 1.818 | 1.484 | 1.533 | 1.560 | 1.450 | 1.563 |
| | AAE | 1.485 | 1.163 | 1.146 | 1.172 | 1.109 | 1.187 |
| | ARE (%) | – | 91.11 | 85.20 | 87.56 | 91.82 | 84.53 |
| | $R^2$ | 0.310 | 0.458 | 0.408 | 0.415 | 0.473 | 0.401 |
| | MBT (S) | 4.416 | 0.698 | 0.083 | 0.639 | 8.359 | 12.715 |
| Roma | RMSE (kWh/m$^2$) | 0.304 | 0.223 | 0.505 | 0.390 | 0.346 | 0.189 |
| | AAE | 0.248 | 0.176 | 0.381 | 0.319 | 0.276 | 0.140 |
| | ARE (%) | 4.95 | 3.54 | 7.70 | 6.38 | 5.54 | 2.80 |
| | $R^2$ | 0.978 | 0.988 | 0.933 | 0.961 | 0.968 | 0.990 |
| | MBT (S) | 7.516 | 0.816 | 0.194 | 3.699 | 11.452 | 36.988 |
| Luxembourg | RMSE (kWh/m$^2$) | 1.188 | 1.175 | 1.115 | 1.081 | 1.114 | 1.083 |
| | AAE | 0.921 | 0.897 | 0.860 | 0.800 | 0.849 | 0.758 |
| | ARE (%) | 57.33 | 53.50 | 42.47 | 41.06 | 43.00 | 28.56 |
| | $R^2$ | 0.690 | 0.704 | 0.736 | 0.756 | 0.722 | 0.750 |
| | MBT (S) | 7.279 | 0.801 | 0.113 | 0.898 | 10.607 | 40.624 |
| Amsterdam | RMSE (kWh/m$^2$) | 0.268 | 0.235 | 0.405 | 0.383 | 0.351 | 1.331 |
| | AAE | 0.207 | 0.157 | 0.259 | 0.255 | 0.213 | 6.312 |
| | ARE (%) | 19.74 | 8.07 | 11.32 | 11.63 | 9.21 | 64.68 |
| | $R^2$ | 0.985 | 0.989 | 0.966 | 0.970 | 0.974 | 0.947 |
| | MBT (S) | 7.327 | 0.68 | 0.114 | 0.853 | 9.549 | 22.66 |
| Budapest | RMSE (kWh/m$^2$) | 1.136 | 1.119 | 1.379 | 1.188 | 1.105 | 1.200 |
| | AAE | 0.903 | 0.886 | 0.967 | 0.862 | 0.846 | 0.822 |
| | ARE (%) | 63.71 | 62.29 | 46.30 | 44.77 | 50.22 | 27.78 |
| | $R^2$ | 0.728 | 0.732 | 0.622 | 0.703 | 0.745 | 0.720 |
| | MBT (S) | 7.344 | 0.705 | 0.12 | 1.989 | 9.148 | 33.432 |
| Vienna | RMSE (kWh/m$^2$) | 1.292 | 1.333 | 1.335 | 1.298 | 1.309 | 1.283 |
| | AAE | 1.041 | 1.057 | 1.061 | 0.996 | 1.010 | 0.968 |
| | ARE (%) | 52.65 | 49.58 | 50.09 | 41.08 | 41.77 | 39.48 |
| | $R^2$ | 0.694 | 0.675 | 0.672 | 0.696 | 0.704 | 0.665 |
| | MBT (S) | 0.156 | 0.666 | 0.163 | 0.709 | 9.634 | 32.777 |
| Dublin | RMSE (kWh/m$^2$) | 1.023 | 0.787 | 0.789 | 0.747 | 0.748 | 0.759 |
| | AAE | 0.768 | 0.674 | 0.569 | 0.564 | 0.549 | 0.491 |
| | ARE (%) | 62.23 | 58.09 | 31.32 | 32.90 | 34.44 | 21.74 |
| | $R^2$ | 0.818 | 0.878 | 0.849 | 0.870 | 0.876 | 0.878 |
| | MBT (S) | 7.448 | 0.672 | 0.126 | 3.243 | 10.606 | 37.941 |
| Sofia | RMSE (kWh/m$^2$) | 0.346 | 0.301 | 0.451 | 0.376 | 0.367 | 0.272 |
| | AAE | 0.263 | 0.228 | 0.314 | 0.251 | 0.242 | 0.200 |
| | ARE (%) | 13.96 | 8.10 | 10.22 | 9.57 | 7.89 | 11.63 |
| | $R^2$ | 0.976 | 0.983 | 0.965 | 0.972 | 0.971 | 0.987 |
| | MBT (S) | 4.553 | 0.733 | 0.132 | 1.773 | 10.008 | 40.941 |
| Warsaw | RMSE (kWh/m$^2$) | 0.270 | 0.202 | 0.390 | 0.349 | 0.303 | 0.264 |
| | AAE | 0.205 | 0.157 | 0.258 | 0.233 | 0.195 | 0.183 |
| | ARE (%) | 17.00 | 11.00 | 10.50 | 10.22 | 8.82 | 12.92 |
| | $R^2$ | 0.981 | 0.990 | 0.959 | 0.968 | 0.982 | 0.980 |

(*continued on next page*)

**Table 5** (*continued*)

| City | Metric | Linear model | Deep Learning | Decision Tree | Random Forest | Gradient Boosted Trees | Support Vector Machine |
|---|---|---|---|---|---|---|---|
| | MBT(S) | 4.396 | 0.74 | 0.109 | 2.881 | 10.086 | 39.448 |
| | RMSE ($kWh/m^2$) | 0.290 | 0.284 | 0.556 | 0.543 | 0.394 | 0.267 |
| Copenhagen | AAE | 0.231 | 0.233 | 0.346 | 0.311 | 0.239 | 0.175 |
| | ARE (%) | 24.81 | 17.11 | 16.67 | 11.21 | 8.54 | 24.78 |
| | $R^2$ | 0.983 | 0.989 | 0.931 | 0.941 | 0.966 | 0.986 |
| | MBT (S) | 4.36 | 0.7 | 0.124 | 2.707 | 10.151 | 46.406 |
| | RMSE ($kWh/m^2$) | 0.306 | 0.271 | 0.456 | 0.498 | 0.391 | 0.210 |
| Stockholm | AAE | 0.235 | 0.193 | 0.277 | 0.300 | 0.231 | 0.128 |
| | ARE (%) | 49.1 | 33.79 | 12.69 | 14.98 | 11.18 | 19.81 |
| | $R^2$ | 0.977 | 0.982 | 0.959 | 0.945 | 0.963 | 0.991 |

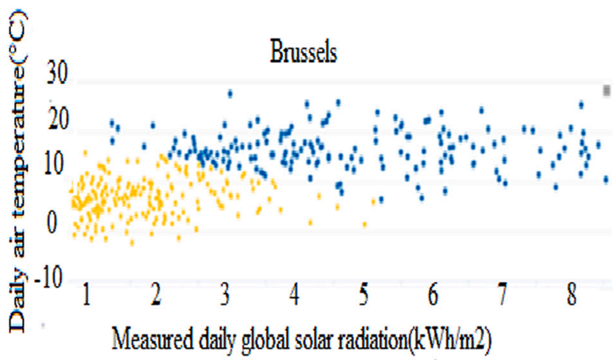


**Fig. 5.** Measured daily global solar radiation function of air temperature in Madrid, Brussels.

size. Finally, the Support Vector Machine model gives an average absolute error of 1.154. This means if you predict H_Gh = 4.04 as a value, the real value is likely between 2.886 and 5.194. The $R^2$ statistic is 0.601. This result reveals that 60.1% of the data fit the regression model. This value is generally considered a moderate effect size.

It is very interesting to notice that in Prague, Linear Model accomplished an average absolute error of 0.147. This means if you predict daily global solar radiation (H_Gh = 0.64) as a value, the real value is likely between 0.493 and 0.787. The $R^2$ statistic is 0.716. This result reveals that 71.6% of the data fit the regression model. Generally, this is a good value. The Deep Learning model accomplished an average absolute error of 0.152. This means if you predict (H_Gh = 0.64) as a value, the real value is likely between 0.488 and 0.792. The $R^2$ statistic is 0.712. This result reveals that 71.2% of the data fit the regression model. This value is generally considered a strong effect size. In addition, the Decision Trees model accomplished an average absolute error of 0.157. This means if you predict (H_Gh = 0.64) as a value, the real value is likely between 0.483 and 0.797. The $R^2$ statistic is 0.690. This result reveals that 69.0% of the data fit the regression model. This value is generally considered a moderate effect size. The Random Forest model accomplished an average absolute error of 0.146. This means if you predict (H_Gh = 0.64) as a value, the real value is likely between 0.494 and 0.786. The $R^2$ statistic is 0.733. This result reveals that 73.3% of the data fit the regression model. Generally, this is a good value. On the other hand, the Gradient Boosted Trees model accomplished an average absolute error of 0.144. This means if you predict daily solar radiation (H_Gh = 0.64) as a value, the real value is likely between 0.496 and 0.784. The $R^2$ statistic is 0.741. This result reveals that 74.1% of the data fit the regression model. Generally, this is a good value. Finally, the Support Vector Machine model accomplished an average absolute error of 0.140. This means if you predict (H_Gh = 0.64) as a value, the real value is likely between 0.5 and 0.78. The $R^2$ statistic is 0.712. This result reveals that 71.2% of the data fit the regression model. Generally, this is a good value. **The** measured daily global radiation function of air temperature are given in fig. (5–6).

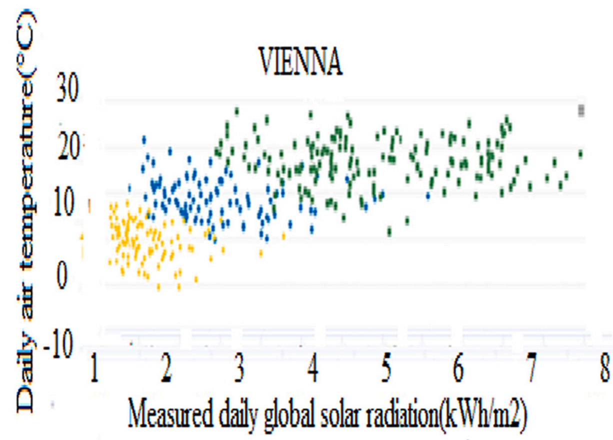


**Fig. 6.** Measured daily global solar radiation function of air temperature in Vienna.

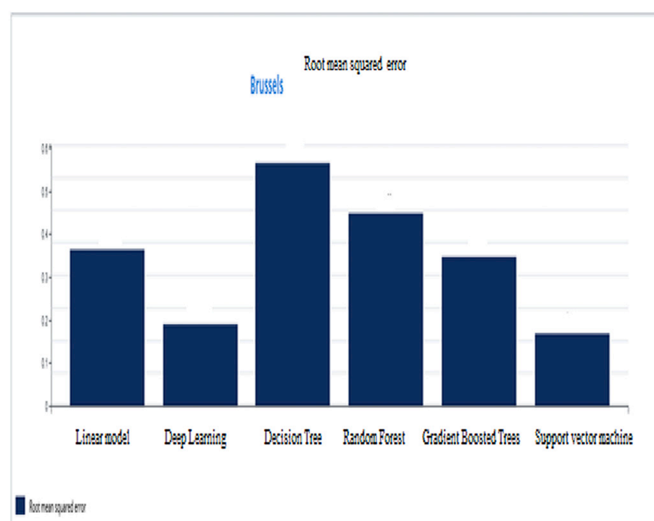


**Fig. 7.** Root mean squared error (RMSE) in some cities according to six machine-learning models.

In each of the figs. (5–6), the different groups are easily identified under the basis of their unique color. Every data is represented by one point. For example group 1, can represent air temperature; group 2, solar radiation; group 3, relative humidity; and group 4, wind speed. In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values. A low standard deviation indicates that the values tend to be close to the mean of the set, while a high standard deviation indicates that the values are spread out over a wider range (Nematchoua et al., 2020). It's very interesting to notice that we got in

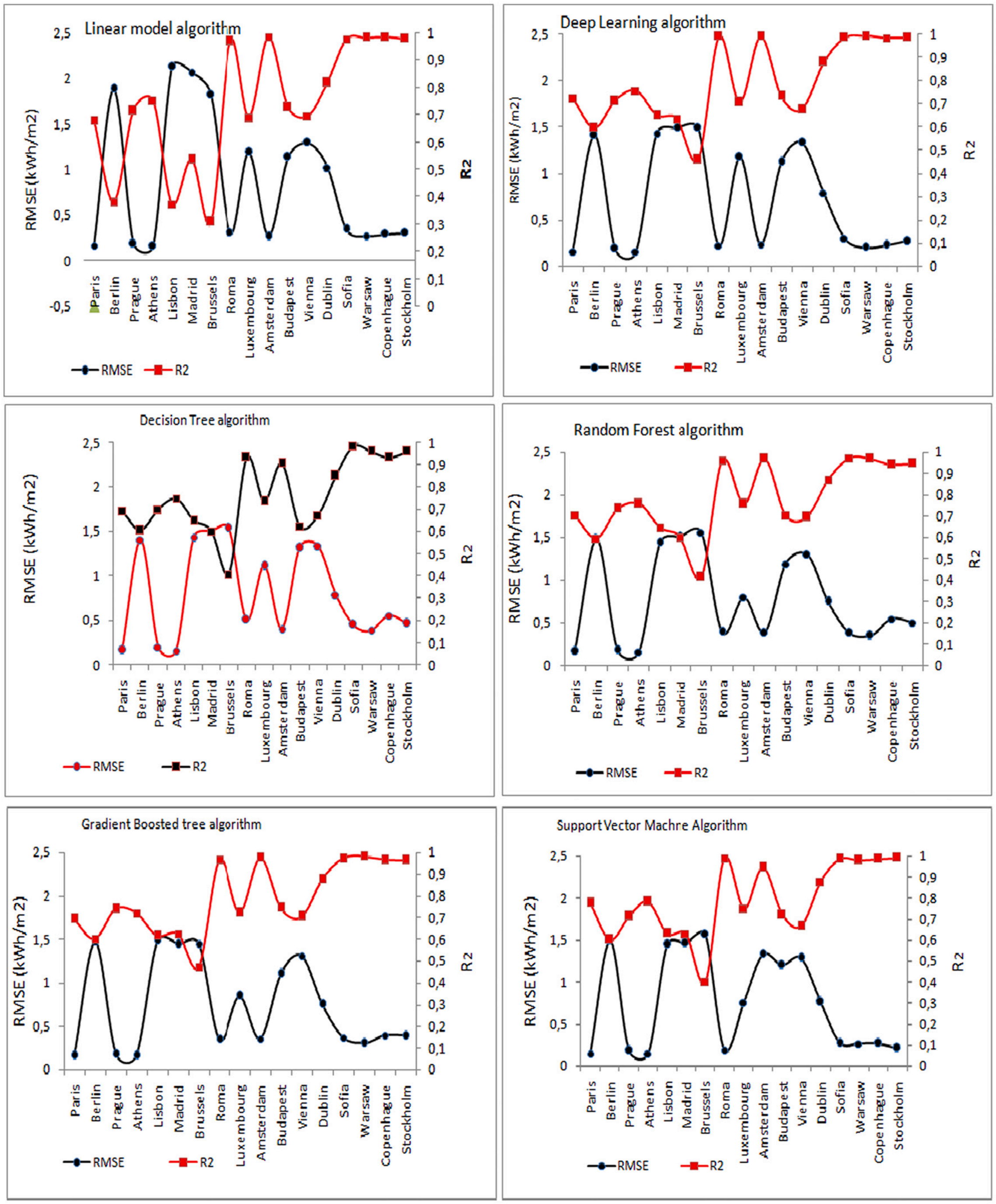


**Fig. 8.** Predictive performance ($R^2$ and RMSE) of the six-machine learning models in several cities in the Europe.

the majority of the case a low standard deviation between the different measured air temperatures and daily global solar radiation for all the cities shown in Figs. 4-6.

Fig. 8 shows the different RMSE in each city according to the machine-learning algorithm. Considering Fig. 7 and Table 5 together, it can be observed that the $R^2$ value of the Deep Learning (DL) algorithm is around 0.898. Among all algorithms, the SVM and DL algorithms give the most successful prediction results in terms of RMSE and MBT. In contrast, the Decision tree algorithm gives the worst prediction results in terms of RMSE in the majority of cities. Fig. 8 gives predictive performance $R^2$ and RMSE of the six machine learning models.

As showed in Fig. 8, the Linear model (RMSE = 0.268, and $R^2$ = 0.985), Deep Learning model (RMSE = 0.235, and $R^2$ = 0.989), Decision Tree model (RMSE = 0.405, and $R^2$ = 0.906), Random Forest model (RMSE = 0.383, and $R^2$ = 0.970) have been the most successful algorithm in terms of statistical metrics in predicting solar radiation data in Amsterdam city. What is more, they are the most unsuccessful in Lisbon (RMSE = 2.126kWh/m$^2$, and $R^2$ = 0.373). It is very important to notice that the Support Vector Machine model has been the most successful algorithm in terms of statistical metrics in predicting solar radiation data (RMSE = 0.189, and $R^2$ = 0.990) in Roma, and also in several other cities. Even if the algorithm with the worst $R^2$ and RMSE values is Decision Tree, this algorithm is also meaningful in terms of AAE. Similar to the Decision Tree algorithm, the Gradient Boosted Tree algorithm had the biggest error magnitude in Budapest.

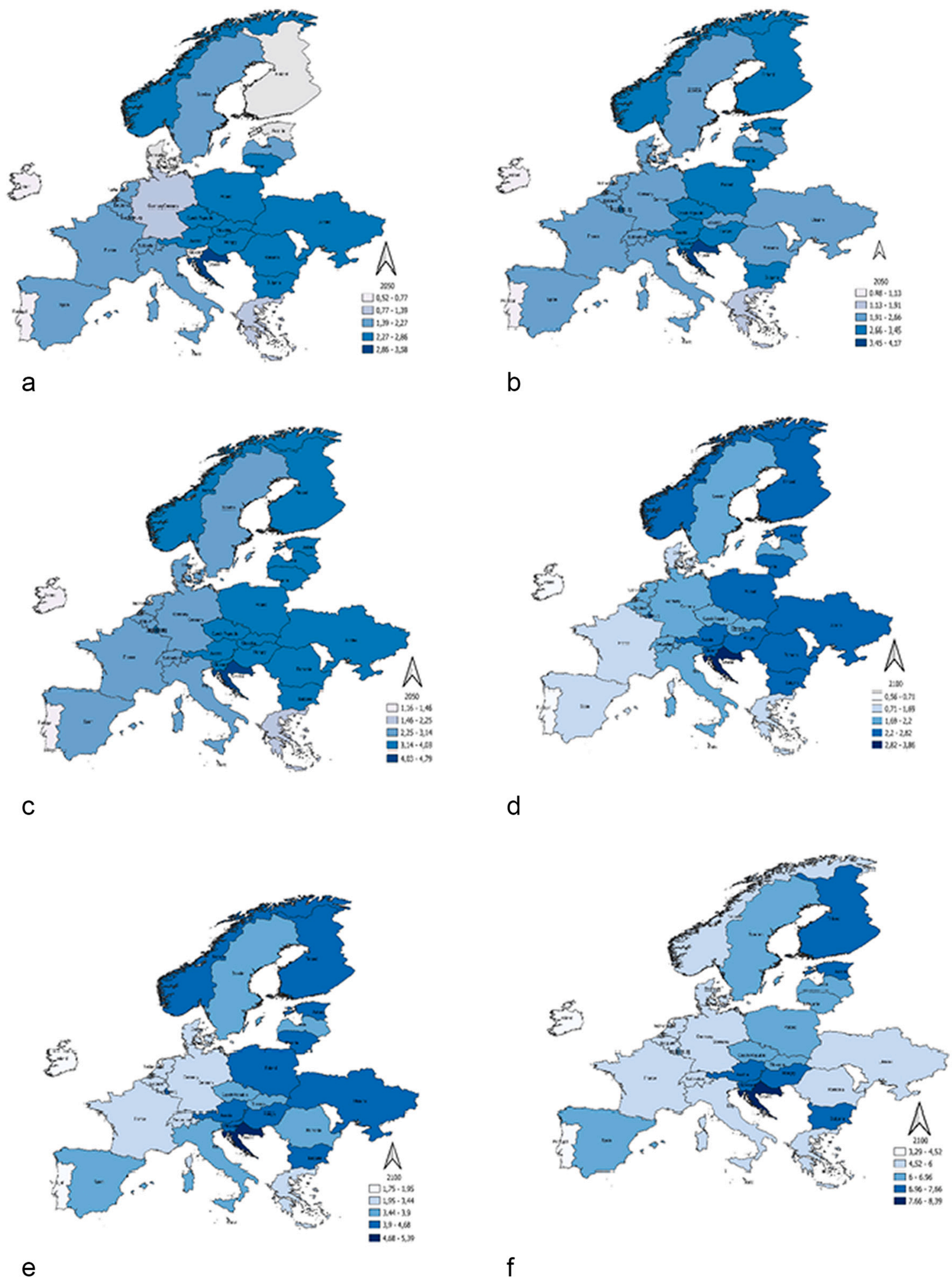


**Fig. 9.** Annually air temperature changes in European Countries according to the RCP scenarios.

### 3.2. Prediction with IPCC scenarios

#### 3.2.1. Variation of air temperature

Fig. 9 shows the mean air temperature increase according to the scenarios RCP 2.6, RCP 4.5 and RCP 8.5 of IPCC by 2050 and 2100 compare to the period 1961–1990.

Analysing the different figures obtained under basic of scenario RCP 2.6, it can be observed deduce that if nothing has done to reduce the evolution of the outdoor climate, the mean outdoor air temperature will increase from 0.52 to 3.58 °C in 2050, and, between 0.56 and 3.86 °C, in 2100, in the 27 countries of EU. The air temperature is expected to increase from 0.52 to 0.77 °C in Portugal and Ireland; from 0.77 to 1.39 °C in Germany and Greece; and, between 2.77 and 2.84 °C, in Poland, Ukraine, Belgium, Austria, Bulgaria etc., in 2050.

According to the scenario RCP 4.5 of the IPCC, in 2050, air temperature is expected to increase between 0.88 and 4.17 °C; in detail,

**Table 6**
Air temperature change and variation of solar radiation.

| IPCC scenarios | Year | Air temperature increase (°C) | Variation of solar radiation ($kWh/m^2$) |
|---|---|---|---|
| Reference | 1961–1990 | 0 | 947–1807 |
| RCP 2.6 | 2050 | 0.52–3.58 | 993–1818 |
| | 2100 | 0.56–3.86 | 1000–1828 |
| RCP 4.5 | 2050 | 0.88–4.17 | 969–1809 |
| | 2100 | 1.75–4.9 | 964–1816 |
| RCP 8.5 | 2050 | 1.16–4.79 | 976–1817 |
| | 2100 | 3.1–5.5 | 981–1837 |

from 1.91 to 1.66 in Spain, France, Belgium, Netherlands etc., between 2.66 and 3.41 °C, in Poland and Bulgaria; and, from 3.45 to 4.17 °C in Croatia. By 2100, it will increase from 1.75 to 4.9 °C, more precisely, between 1.95 and 3.44 °C in France, Belgium, Germany, Switzerland etc., and from 3.9 to 4.6 °C in Norway, Finland, Hungry, and Estonia etc.

Regarding scenario 8.5 of the IPCC, it is important to notice that the air temperature is expected to increase by 2050 between 1.16 and 4.79 °C, either from 3.14 and 4.03 °C in Ukraine, Romania, and Bulgaria. In addition, in 2100, if nothing has been done to reduce the evolution of the outdoor climate, the mean outdoor air temperature will increase from 3.1 to 5.5 °C. In detail, from 3.9 to 4.4 °C in France, Germany and Belgium; from 4.4 to 4.6 °C in Sweden, Poland and Lithuania; and between 4.6 and 4.9 °C in Hungry, Austria, Bulgaria etc.

All these results show that the air temperature increase also depends on the geo-location of the country.

It is very interesting to notice that air temperature will be from 0.55 to 1.02 °C higher in coastal countries than continental countries in the next decade in European Union countries.

### 3.2.2. Variation of global solar radiation

As shown in the last column in Table 5, the variation of the global solar radiation in the 27 European countries is remarkable. Indeed, in 2050, it will vary from 993 to 1818 $kWh/m^2$, according to scenario RCP2.6; from 969 to 1809 $kWh/m^2$ (RCP4.5); between 976 and 1817 $kWh/m^2$ (RCP8.5). However, in 2100, the global solar radiation is expected to be between 1000 and 1828 $kWh/m^2$, 964–1816 $kWh/m^2$, and 981–1837 $kWh/m^2$, according to the scenarios RCP2.6, RCP 4.5, and RCP 8.5, respectively. The detailed results are grouped in Table 6.

The Europe countries can be classified according to the global solar radiation scale between 2050 and 2100:

(1) 993–1029 $kWh/m^2$, case of Ireland, Norway, Sweden and Finland;

(2) 1024–1140 $kWh/m^2$, case of Belgium, Germany, Luxembourg, Ukraine etc.

(3) 1188–1607 $kWh/m^2$, case of Hungary, Romania, Italy, Bulgaria etc.

(4) 1583–1818 $kWh/m^2$, case of Spain, Portugal, Greece.

It noticed that all the input data strongly impact predicted radiation. Despite this, good correlations are obtained between measured and predicted daily global solar radiation. Fig. 10 shows the Global irradiation predicted in 2050 and 2100 using three current IPCC scenarios in European countries.

The solar radiation received on a surface, therefore, varies over time depending on the position of the Sun, and other meteorological parameters, however, this radiation is inexhaustible.

The essential source of energy for the Earth's surface is the flow of solar energy. If the flux received at the surface of the Earth in a given location varies considerably, especially seasonally, the flux radiated by the Sun is relatively constant. However, signs of variations in the activity of the Sun are known for a very long time, and the hypothesis that this solar activity can affect our climate is old.

## 4. Discussions and comparisons

### 4.1. Analyse and comparison of prediction algorithms

Table 6 shows the comparisons between some results from the literature study and this research.

It is very interesting to notice that the number of preferred metrics in determining the prediction success of all the algorithms is limited. Indeed, the majority of researches conducted regarding the prediction of global solar radiation used algorithms coming from the same classes and categories. Therefore, all the statistical metrics generally showed close results to every other for a different type of algorithm (Chen et al., 2013). Often, the limited number of metrics makes it very difficult to evaluate the prediction success of the models. For example, in 2011, Moreno et al. (Moreno et al., 2011) analysed the daily global solar radiation data. These authors found one of the highest $R^2$ values with 0.86 in their study. This value is one of the best ones in Table 6. However, the RMSE value was calculated as 0.495 $kWh/m^2$ in the relevant research (Chen et al., 2013). This value is also identified as the worst one in Table 7. This raises a question in the minds and sometimes makes it difficult to select the best simulation model. In this study, four metrics are applied, which orient us to discuss and select the best predict model.

Another important point noticed by analysing Table 7 is that there is not any algorithm still showing the best results for all cities. Even making the different predictions with the same data type, it is often noticed that the algorithms give the best results city to city. Therefore, there may have a difference same among the metrics having the same algorithms presenting the best results for different cities. For example, in 2016, Mehdizadeh et al. (Mehdizadeh et al., 2016) achieved the best results in the prediction of daily global solar radiation using the ANN algorithm. In the relevant research, the good value of RMSE was calculated as 1.850 (Mehdizadeh et al., 2016). In another research in which achieving the best prediction results with ANN, Antonopoulos et al. found the highest RMSE value as 3.166 (Antonopoulos et al., 2019).

Some reasons causing this difference may be due to the local climate, geographical differences, missing data, input variables, feature selection, dataset size etc.

As explained in the previous sections, in this study, the best prediction results are achieved with the SVM algorithm in the majority of capitals located in the European Union. Table 5 showed some studies, where the results obtained with the SVM algorithm are considered. This research has mostly presented one of the best metric results in comparison with those of other studies (Chen et al., 2013; Mohammadi et al., 2015a; Quej et al., 2017b). The main reason behind it may be due to dataset diversity, geographical advantages and almost no missing data in the dataset. Based on the previous results, we can deduce that the best machine-learning model allows getting the prediction data with good accuracy.

### 4.2. Analysis and comparison of air temperature and global solar radiation in the future

The results of this research show that it will be possible in European countries in 2100, to see a temperature change from 0.5 to 3.8 °C (according to the scenario RCP 2.6); from 3.1 to 5.5 °C (scenario RCP 8.5), compare to the period 1960–1990. These temperature increase could impacted on global solar radiation, which should to be between 993 and 1837 $kWh/m^2$. These results enormously depend on the data category used during the prediction and model. The prediction of IPCC showed a air temperature increase of 1.5–2.5 °C (according to the scenario 2.6), and, from 3.5 and 5.5 °C (regarding the scenario 8.5) in 2100 (see subsection 2.5.3.). The results of this study are consistent with those proposed by the IPCC. Others research presented the similar results, but with different period of comparison (An et al., 2020; An et al., 2021; Babatunde et al., 2020a; Babatunde et al., 2020b; Modeste et al., 2020;

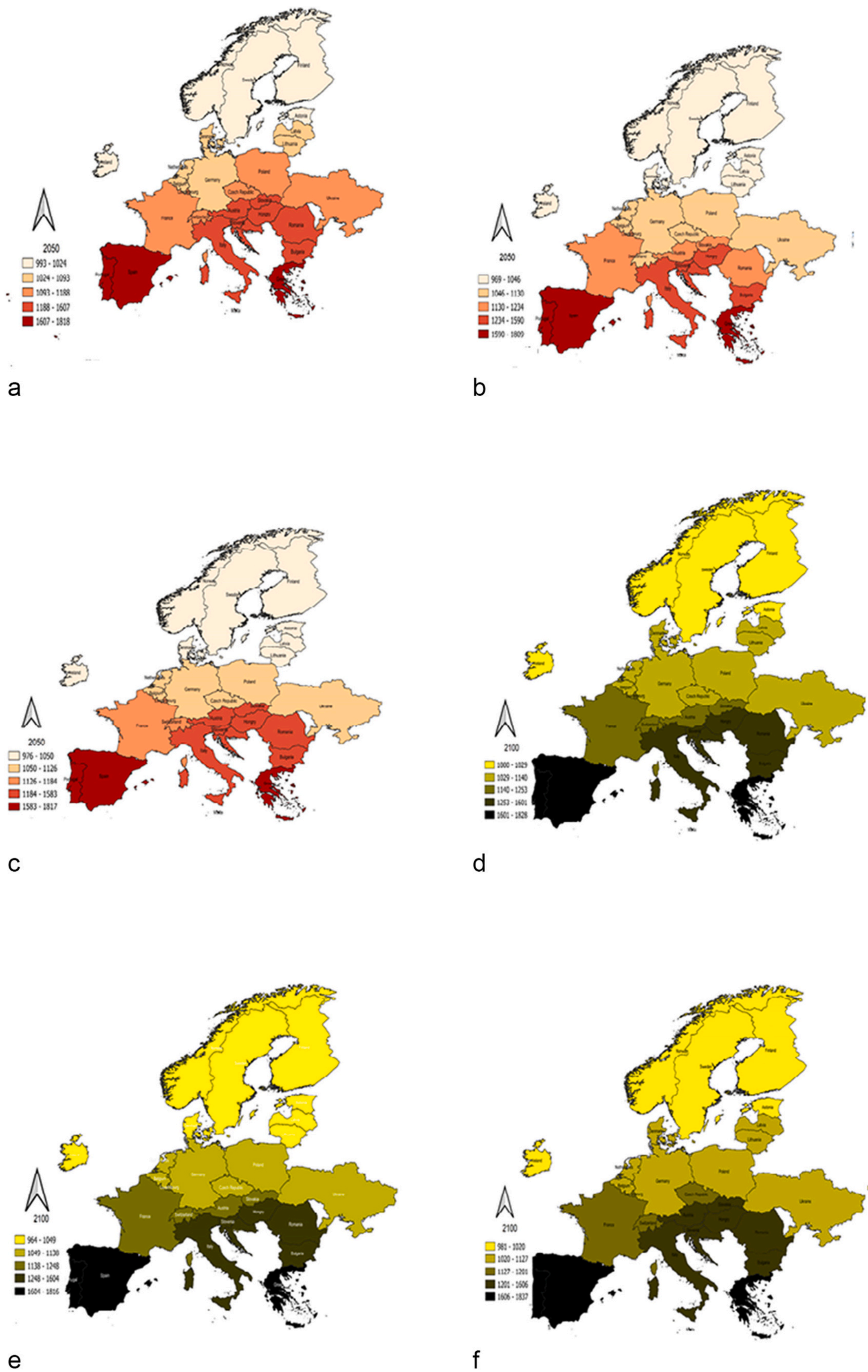


**Fig. 10.** Global irradiation in European countries in 2050 and 2100 according to IPCC scenarios.

**Table 7**
Comparison of the present study with the literature researches in the prediction of daily global solar radiation (NA = Not applicable).

| Ref. | Prediction models | Best model | Evaluation metrics | | | |
|---|---|---|---|---|---|---|
| | | | RMSE (kWh/m$^2$) | AAE | ARE (%) | $R^2$ |
| (Chen et al., 2013) | SVM, Empirical | SVM | 0.495 | NA | NA | NA |
| (Moreno et al., 2011) | BC,ANN,KNN | ANN | 0.875 | NA | NA | 0.86 |
| (Mohammadi et al., 2015a) | SVM,WT, ANN,ARMA | SVM | 0.394 | NA | NA | 0.91 |
| (Wang et al., 2016) | MLP,RBF, GRNN | MLP | 0.537 | NA | NA | 0.86 |
| (Mohammadi et al., 2015b) | SVR | SVR | 0.555 | NA | NA | 0.91 |
| (Wang et al., 2017) | ANFIS, Empirical MSTree | ANFIS | 0.573 | NA | NA | 0.91 |
| (Aji et al., 2018) | SVR | SVR | NA | NA | NA | 0.98 |
| (Quej et al., 2017b) | ANFIS,SVM, ANN | SVM | 0.714 | NA | NA | 0.69 |
| (Mehdizadeh et al., 2016) | GEP,ANN, ANFIS | ANN | 0.512 | NA | NA | 0.93 |
| (Feng et al., 2019) | MEA-ANN, ANN, RF, WNN, Empirical | MEA-ANN | 0.758 | NA | NA | 0.88 |
| (Jahani and Mohammadi, 2019) | ANN, Empirica | ANN | 0.513 | NA | NA | 0.92 |
| (Antonopoulos et al., 2019) | ANN, MLR | ANN | 0.876 | NA | NA | 0.88 |
| (Shamshirband et al., 2016) | ANN-ARX | ANN-ARX | 0.479 | NA | NA | 0.87 |
| (Shamshirband et al., 2015) | KELM | KELM | 0.558 | NA | NA | 0.82 |
| (Agbulut et al., 2021) | ANN,SVM, DL,k-NN | ANN | 0.597 | NA | NA | 0.93 |
| This study | LM,DL,DT, GB, SVM,RF | SVM | 0.708 | 0.176 | 38.39 | 0.80 |

Morkovkin et al., 2020; Nematchoua et al., 2019b; Nematchoua et al., 2019f; Nematchoua et al., 2020; Nematchoua et al., 2021a; Nematchoua et al., 2021b). The Fig. 11 and 12 show the variation of global solar radiation in some cities in Europe.

At the same time, Table 8 gives some statistical analysis of data represented in the previous figures.

The RCPs scenarios (2.6; 4.5; 6.0; 8.5 etc.) of the IPCC are considered to be the most reliable in the prediction of climate data.

Between (1960–1990) and 2050, minimum solar radiation will vary from 946.98 to 975.73 kWh/m$^2$; whereas, mean solar radiation will varying from 1189.00 to 1247.96 kWh/m$^2$ (scenario RCP8.5). This variation may be due to the effect of humans on nature. These results confirm research conducted by de Larsen et al. (Larsen et al., 2020) in 2020, and results of other studies detailed (Maillo et al., n.d.; Saikia et al., n.d.; Erdal and Aytug, 2016; Min et al., 2003; Nordhaug et al., 2018; Pardis et al., 2019).

## 5. Conclusion

This study evaluates the performance of six different machine learning algorithms (LM, DL, DT, GB, SVM, and RF) in the prediction of daily global solar radiation. The research considers various input data (Relative Humidity, air temperature, wind speed, and global solar radiation) from the twenty-seven different stations (Paris, Brussels, Roma, Amsterdam, Sofia etc.) in Europe. To evaluate the performance of the machine learning algorithms, four metrics ($R,^2$ RMSE, AAE, and ARE) are discussed in this research. Some important conclusions can be drawn based on the present result.

(a) Under the basis of the prediction results in terms of $R,^2$ it can be deduced that algorithms used in all the cities have shown very successful results. Indeed, $R^2$ values of the algorithms in this research varied from 53.7% to 98.6% depending on the cities, excepted in Berlin (38.2%), Lisbon (37.3%), and Brussels (31%).

(b) Regarding all statistical metrics together for 27 cities in Europe, the best results are achieved within Amsterdam city.

(c) When all cities and algorithms are evaluated in terms of RMSE, it is seen that the majority of values obtained are nearly zero. That is, mean that the majority of prediction results can be categorized as "reasonable prediction", or "good prediction".

(d) It is very interesting to notice that, during the general evaluation it was found that DL and SVM algorithms give very close results, considering the $R^2$ and RMSE metrics. Therefore, different parameters should be taken into account to discuss and select the best of these algorithms.

(e) One of the most important parameters that distinguish The DL

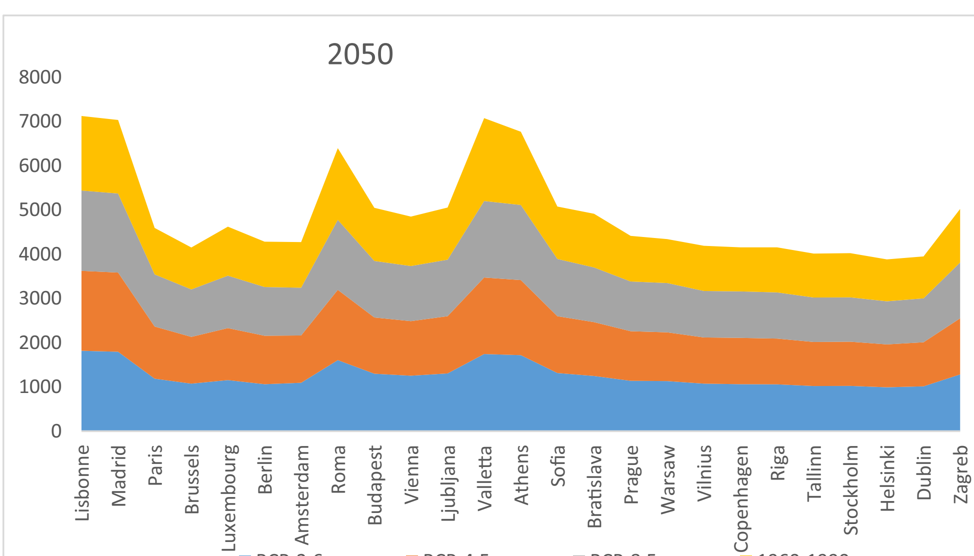


**Fig. 11.** Variation of solar radiation in some cities in European in 2050 (in kWh/m$^2$).

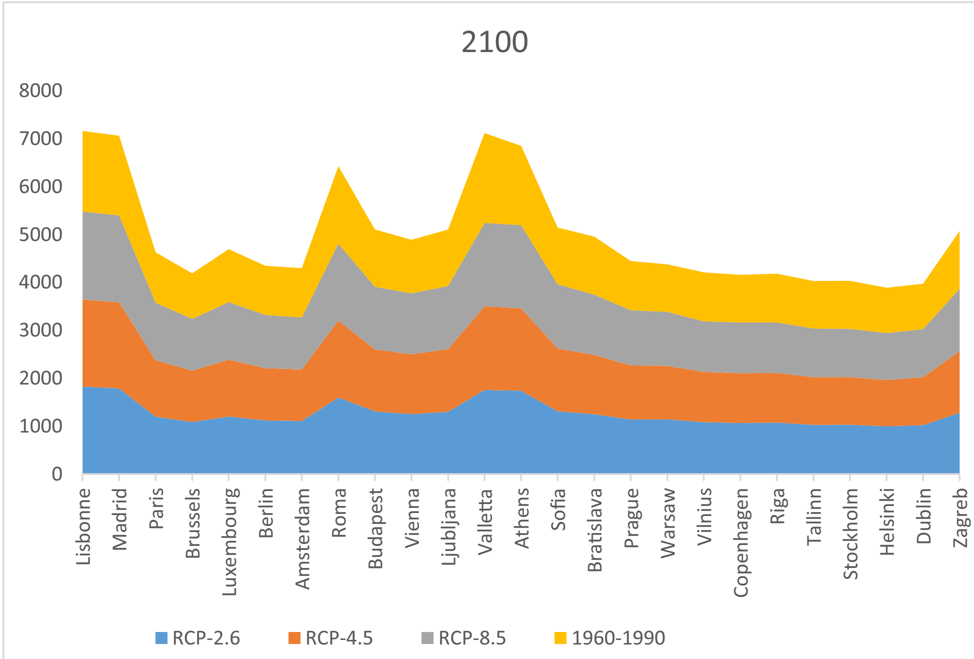


**Fig. 12.** Variation of solar radiation in some cities in the European in 2100 (in kWh/m$^2$).

**Table 8**
Statistical analysis of the future global solar radiation (in kWh/m$^2$).

| Year | Statistical parameter | RCP-2.6 | RCP-4.5 | RCP-8.5 |
|---|---|---|---|---|
| 2050 | Standard deviation | 204.32 | 203.76 | 200.188 |
| | Minimum | 992.8 | 969.33 | 975.73 |
| | Maximum | 1817.93 | 1809.01 | 1816.99 |
| | Mean | 1259.55 | 1243.49 | 1247.96 |
| 2100 | Standard deviation | 199.91 | 209.09 | 208.66 |
| | Minimum | 999.50 | 963.95 | 980.62 |
| | Maximum | 1828.21 | 1815.78 | 1836.72 |
| | Mean | 1269.01 | 1252.70 | 1266.5 |

and SVM algorithms is MBT. Even if these two algorithms are very close to each other particularly in the majority of cities, the SVM algorithm showed higher MBT than DL. However, analysing all the error magnitudes of the observations randomly determined in this research, we can deduce that the error magnitudes in the use of the SVM algorithm are very low in comparison with those of other algorithms used.

(f) Repetitive heat waves spread over long periods are expected between 2050 and 2100 in Madrid, Athens and Lisbon.

These results can help the European governments for better decision-making regarding the integration of the sun as one of the main sources of electricity in the future.

## Author contributions

Conceptualization, M.K·N.; methodology, M.K.N., M.A., and J.A.O., formal analysis, M.K.N.,; investigation, M.K.N., data curation, M.K.N.; writing—original draft preparation, M.K.N., and J.A.O., writing—review and editing, M.K.N., J.A.O., and M.A.; Supervision, M.K.N ., All authors have read and agreed to the published version of the manuscript.

## Funding

The work presented in this paper has not received external funding.

## Institutional Review Board Statement

Not applicable.

## Informed Consent Statement

Not applicable.

## Data Availability Statement

Data available under request of reader.

## Declaration of Competing Interest

The authors declare no conflict of interest.

## References

Agbulut, Ümit, Gürel, Ali Etem, Biçen, Yunus, 2021. Prediction of daily global solar radiation using different machine learning algorithms: evaluation and comparison. Renew. Sust. Energ. Rev. 135, 110114.
Aji, P., Wakamori, K., Mineno, H., 2018. Highly accurate daily solar radiation forecasting using SW-SVR for hybrid power plant in Indonesia. In: 2018 4th international conference on nano electronics research and education (ICNERE). IEEE, pp. 1–5.
An, J., Mikhaylov, A., Richter, U.H., 2020. Trade war effects: evidence from sectors of energy and resources in Africa. Heliyon 6, e05693. https://doi.org/10.1016/j.heliyon.2020.e05693.
An, J., Mikhaylov, A., Jung, S.-U., 2021. A linear programming approach for robust network revenue management in the airline industry. J. Air Transp. Manag. 91 (3), 101979 https://doi.org/10.1016/j.jairtraman.2020.101979.
Antonopoulos, V.Z., Papamichail, D.M., Aschonitis, V.G., Antonopoulos, A.V., 2019. Solar radiation estimation methods using ANN and empirical models. Comput. Electron. Agric. 160, 160–167.
Babatunde, Olubayo M., Munda, Josiah L., Hamam, Yskandar, 2020a. Energies 13, 2488. https://doi.org/10.3390/en13102488.
Babatunde, O.M., et al., 2020b. Energies 13, 2488.
Bakirci, K., 2009. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. Energy 34 (4), 485–501.
Balog, I., Podrascanin, Z., Spinelli, F., Caputo, G., Siviero, R., Benedetti, A., 2019. Hourly forecast of solar radiation up to 48h with two runs of weather research forecast model over Italy. In: AIP Conference Proceedings, 2126. AIP Publishing LLC, p. 190004 (No. 1).
Bayrakçı, H.C., Demircan, C., Keçebaş, A., 2018. The development of empirical models for estimating global solar radiation on horizontal surface: a case study. Renew. Sust. Energ. Rev. 81, 2771–2782.

Beer, C., Reichstein, M., Tomelleri, E., Ciais, P., Jung, M., Carvalhais, N., et al., 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate. Science 329, 834–838. https://doi.org/10.1126/science.11 84984.
Bontempi, Gianluca, 2021. Statistical foundations of machine learning. In: Handbook, Second edition, p. 4(377).
Chen, J.L., Liu, H.B., Wu, W., Xie, D.T., 2011. Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study. Renew. Energy 36, 413–420. https://doi.org/10.1016/j.renene.2010.06.024.
Chen, J.L., Li, G.S., Wu, S.J., 2013. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy Convers. Manag. 75, 311–318.
CNRS, 2020. Le rayonnement solaire. https://sagascience.cnrs.fr/ (Accessed on 24 July 2021).
Erdal, Tasci, Aytug, Onan, 2016. Investigation of the Effect of K-Nearest Neighbour Algorithm Parameters on Classification Performance. In: Turkish akademik bilis¸im.
Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X., Xiang, Y., 2018a. New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China. Energy Convers. Manag. 156, 618–625.
Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., et al., 2018b. 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. 164, 102–111. https://doi.org/10.1016/j.enconman.2018.02.087.
Fan, J., Wang, X., Wu, L., Zhang, F., Bai, H., Lu, X., Xiang, Y., 2018c. New combined models for estimating daily global solar radiation based on sunshine duration in humid regions: a case study in South China. Energy Convers. Manag. 156, 618–625.
Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L., Cui, N., 2019. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. Energy Convers. Manag. 198, 111780.
Freedman, David A., 2009a. Statistical models: theory and practice. In: A Simple Regression Equation has on the Right Hand Side an Intercept and an Explanatory Variable with a Slope Coefficient. A Multiple Regression e Right Hand Side, each with its Own Slope Coefficient. Cambridge University Press, p. 26.
Freedman, David A., 2009b. Statistical models: theory and practice. In: A Simple Regression Equation has on the Right Hand Side an Intercept and an Explanatory Variable with a Slope Coefficient. Cambridge University Press, p. 26.
Garland, F.C., Garland, C.F., Gorham, E.D., Young, J.F., 1990. Geographic variation in breast cancer mortality in the United States: a hypothesis involving exposure to solar radiation. Prev. Med. (Baltim). 19, 614–622. https://doi.org/10.1016/0091-7435(90)90058-R.
Guher, Abdurrahman Burak, Tasdemir, Sakir, Yaniktepe, Bulent, 2020. Effective Estimation of Hourly Global Solar Radiation Using Machine Learning Algorithms. Int. J. Photoenergy. https://doi.org/10.1155/2020/8843620. Article ID 8843620, 26 pages.
Gürel, A.E., Agbulut, Ü., Biçen, Y., 2020. Assessment of machine learning, time series, response surface methodology and empirical models in prediction of global solar radiation. J. Clean. Prod. 122353.
Hastie, T., Tibshirani, R., Friedman, J.H., 2009. "10. Boosting and Additive Trees". The Elements of Statistical Learning, 2nd ed. Springer, New York, pp. 337–384 (ISBN 978–0–387-84857-0. Archived from the original on 2009-11-10).
Ho, Tin Kam, 1995. Random Decision Forests (PDF). In: Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16 August 1995, pp. 278–282.
Ho, T.K., 2021. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844. https://doi.org/10.1109/34.709601.
https://en.wikipedia.org/wiki/Representative_Concentration_Pathway. Consulted on July 06, 2021.
Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., Zeng, Z., 2001. Solar radiation prediction using different machine Learning algorithms and implications for extreme climate events. Front. Earth Sci. 9, 596860 https://doi.org/10.3389/feart.2021.596860.
Huang, L., Kang, J., Wan, M., Fang, L., Zhang, C., Zeng, Z., 2021. Solar radiation prediction using different machine Learning algorithms and implications for extreme climate events. Front. Earth Sci. 9, 596860 https://doi.org/10.3389/feart.2021.596860.
Jahani, B., Mohammadi, B., 2019. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. Theor. Appl. Climatol. 137 (1–2), 1257–1269.
Jiang, Y., Long, H., Zhang, Z., Song, Z., 2017. Day-ahead prediction of bihourly solar radiance with a Markov switch approach. IEEE Transact. Sustain. Energy 8 (4), 1536–1547.
Kamiński, B., Jakubczyk, M., Szufel, P., 2017. A framework for sensitivity analysis of decision trees. CEJOR 26 (1), 135–159. https://doi.org/10.1007/s10100-017-0479-6. PMC 5767274. PMID 29375266.
Karimi, K., Hamilton, H.J., 2011. "generation and interpretation of temporal decision rules", international journal of computer information systems and industrial management applications. Volume 3.
Larsen, M.A.D., Petrovic, S., Radoszynski, A.M., McKenna, R., Balyk, O., 2020. Climate change impacts on trends and extremes in future heating and cooling demands over Europe. Energy & Build. 226, 110397.
Liu, Y., Zhou, Y., Chen, Y., Wang, D., Wang, Y., Zhu, Y., 2020. Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: a case study in China. Renew. Energy 146, 1101–1112.
Long, H., Zhang, Z., Su, Y., 2014. Analysis of daily solar power prediction with data-driven approaches. Appl. Energy 126, 29–37.
Machine Learning, 2019. https://machinelearningmastery.com/linear-regression-for-machine-learning/. Available online in 2019. (Accessed on 24 July 2021).
Maillo, J., Luengo, J., García, S., Herrera, F., 2017. Exact fuzzy k-nearest neighbor classification for big datasets. In: IEEE International Conference on Fuzzy Systems. FUZZ-IEEE. IEEE, p. 2017.
Marzo, A., Trigo-Gonzalez, M., Alonso-Montesinos, J., Martínez-Durban, M., Lopez, G., Ferrada, P., Batlles, F.J., 2017. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. Renew. Energy 113, 303–311.
Meenal, R., Selvakumar, A.I., 2018. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. Renew. Energy 121, 324–343.
Mehdizadeh, S., Behmanesh, J., Khalili, K., 2016. Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. J Atmos Sol Terr Phys 146, 215–227.
Meteo Belgium. https://www.meteo.be/fr/climat/climat-de-la-belgique/atlas-climatique/cartes-climatiques/rayonnement-solaire/a-propos (Accessed on 24 July 2021).
Meteonorm. https://meteonorm.com/en/meteonorm-version-8 (Accessed on 24 July 2021).
Meteorological data from ground stations. https://marswiki.jrc.ec.europa.eu/agri4castwiki/index.php/Meteorological_data_from_ground_stations (Accessed on 24 July 2021).
Min, Yoon, Yeboon, Yun, Hirotaka, Nakayama, 2003. A role of total margin in support vector machines. In: Proceedings of the International Joint Conference on Neural Networks. IEEE, p. 2003.
Modeste, K.N., Jose, A.O., Cinzia, B., Esther, O., Donghyun, R., Paola, R., Sigrid, R., 2020. Comparative analysis of bioclimatic zones, energy consumption, $CO_2$ emission, and life cycle cost of residential and commercial buildings located in a tropical region, A case study of the big island of Madagascar. Energy. 202, 117754.
Mohammadi, K., Shamshirband, S., Tong, C.W., Arif, M., Petkovi'c, D., Ch, S., 2015a. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers. Manag. 92, 162–171.
Mohammadi, K., Shamshirband, S., Anisi, M.H., Alam, K.A., Petkovi'c D., 2015b. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Convers. Manag. 91, 433–441.
Moreno, A., Gilabert, M.A., Martínez, B., 2011. Mapping daily global solar irradiation over Spain: a comparative study of selected approaches. Sol. Energy 85 (9), 2072–2084.
Morgan. Nelson., 2011. Deep and wide: multiple layers in automatic speech recognition. IEEE Trans. Audio Speech Lang. Process. 20 (1), 7–13.
Morkovkin, D.E., Gibadullin, A.A., Kolosova, E.V., Semkina, N.S., Fasehzoda, I.S., 2020. Modern transformation of the production base in the conditions of Industry 4.0: problems and prospects. J. Phys. Conf. Ser. 1515, 032014.
Najafabadi, Maryam M., Villanustre, Flavio, Khoshgoftaar, Taghi M., 2015. Naeem Seliya, Randall Wald and Edin Muharemagic. Deep learning applications and challenges in big data analytics. J. Big Data 2 (1), 1.
Neelamegam, Premalatha, Amirtham, Valan Arasu, 2016. Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. J. Appl. Res. Technol. 14, 206–214.
Nematchoua, M.K., Jacques, T., Reiter, S., 2019b. Statistical life cycle assessment of residential buildings in a temperate climate of northern part of Europe. J. Clean. Prod. 229, 621–631.
Nematchoua, M.K., Orosa, J.A., Reiter, S., 2019f. Climate change: variabilities, vulnerabilities and adaptation analysis - A case of seven cities located in seven countries of Central Africa. Urban Clim. 29, 100486.
Nematchoua, M.K., Noelson, J.C.V., Saadi, I., Sela, J.B., Reiter, S., 2020. Application of phase change materials, thermal insulation, and external shading for thermal comfort improvement and cooling energy demand reduction in an office building under different coastal tropical climates. Sol. Energy 207, 458–470 this link is disabled.
Nematchoua, Modeste Kameni, Nishimwe, Antoinette Marie-Reine, Reiter, Sigrid, 2021a. Towards nearly zero-energy residential neighbourhoods in the European Union: A case study. Renew. Sust. Energ. Rev. 135, 110198.
Nematchoua, M.K., Orosa, J.A., Ricciardi, P., Obonyo, E., Sambatra, E.J.R., Reiter, S., 2021b. Transition to zero energy and low carbon emission in residential buildings located in tropical and temperate climates. Energies 14, 4253. https://doi.org/10.3390/en14144253.
Nordhaug, Myhre Jonas, et al., 2018. Robust clustering using a kNN mode seeking ensemble. Pattern Recogn. 76, 491–505.
Pardis, Birzhandi, Kim, Kyung Tae, Lee, Byungjun, Youn, Hee Yong, 2019. Reduction of training data using parallel hyperplane for support vector machine. Appl. Artif. Intell. 33 (6), 497–516.
Persson, C., Bacher, P., Shiga, T., Madsen, H., 2017. Multi-site solar power forecasting using gradient boosted regression trees. Sol. Energy 150, 423–436. https://doi.org/10.1016/j.solener.2017.04.066.
Pielke, J.R., Roger, April 10, 2021. Most plausible 2005–2040 emissions scenarios project less than 2.5 degrees C or warming by 2100. osf.io. https://doi.org/10.31235/osf.io/m4fdu. Retrieved 2021-04-26.
Piryonesi, S. Madeh, El-Diraby, Tamer E., 2020. Data analytics in asset management: cost-effective prediction of the pavement condition index. J. Infrastruct. Syst. 26 (1), 04019036. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000512. ISSN 1943-555X.
Piryonesi, S. Madeh, El-Diraby, Tamer E., 2021. Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration

modeling. J. Infrastruct. Syst. 27 (2), 04021005. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000602. ISSN 1076–0342.
Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L., 2017a. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Sol Terr Phys 155, 62–70.
Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L., 2017b. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Sol Terr Phys 155, 62–70.
Quinlan, R., 1983. Learning efficient classification procedures. In: Michalski, Carbonell, Mitchell (Eds.), Machine Learning: an artificial intelligence approach. Morgan Kaufmann, pp. 463–482. https://doi.org/10.1007/978-3-662-12405-5_15.
Quinlan, J.R., 1987a. Simplifying decision trees. Int. J. Man-Mac. Stud. 27 (3), 221–234.
Quinlan, J.R., 1987b. Simplifying decision trees. Int. J. Man-Mach. Stud. 27 (3), 221–234. CiteSeerX 10.1.1.18.4267. https://doi.org/10.1016/S0020-7373(87)80053-6.
Rehman, S., 1998. Solar radiation over Saudi Arabia and comparisons with empirical models. Energy 23 (12), 1077–1082.
Rencher, Alvin C., Christensen, William F., 2012. Chapter 10, multivariate regression – Section 10.1, introduction. In: Methods of Multivariate Analysis, Wiley Series in Probability and Statistics, , 3rd ed.709. John Wiley & Sons, p. 19. ISBN 9781118391679.
Saikia, J., Yin, S., Jiang, Z., Seok, M., Jae-sun, S., 2019. K-nearest neighbor hardware accelerator using inmemory computing SRAM. In: IEEE/ACM international symposium on low power electronics and design (ISLPED). IEEE, p. 2019.
Seal, Hilary L., 1967. The historical development of the gauss linear model. Biometrika. 54 (1/2), 1–24. https://doi.org/10.1093/biomet/54.1-2.1. JSTOR 2333849.
Shamshirband, S., Mohammadi, K., Chen, H.L., Samy, G.N., Petkovi'c D, Ma C., 2015. Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: a case study for Iran. J Atmos Sol Terr Phys 134, 109–117.
Shamshirband, S., Mohammadi, K., Piri, J., Petkovi'c, D., Karim, A., 2016. Hybrid autoregressive neural network model for estimating global solar radiation in Bandar Abbas, Iran. Environ. Earth Sci. 75 (2), 172.
Shaohui, M., Ji, J., Hou, J., Li, X., Du, Q., 2017. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. IEEE Trans Geosci Rem Sens 55 (8), 4520–4533.
Station météo. https://www.meteobelgique.be/index.php?option=com_static&staticfile=realtime-station_list.php&Itemid=69 (Accessed on 24 July 2021).
Sullivan, J.P., Carver, M.R., Kippen, R.M., Friedel, R.H., Reeves, G.D., Henderson, M.G., 2017. Energetic Particle Data From the Global Positioning System Constellation, p. 10. https://doi.org/10.1002/2017SW001604.
Sun, H., Gui, D., Yan, B., Liu, Y., Liao, W., Zhu, Y., et al., 2016. Assessing the potential of random forest method for estimating solar radiation using air pollution index. Energy Convers. Manag. 119, 121–129. https://doi.org/10.1016/j.enconman.2016.04.051.
Talib, Manar Abu, Majzoub, Sohaib, Nasir, Qassim, Jamal, Dina, 2016. DLAU: a scalable deep learning accelerator unit on FPGA. IEEE Trans Comput Aided Des Integrated Circ Syst 36 (3), 513–517.
TSMS, January 7, 2020. Turkish State Meteorological Service. https://mgm.gov.tr/eng/ (Accessed on 24 July 2021).
Wagner, Harvey M., 1 September 1975. Principles of Operations Research: With Applications to Managerial Decisions, 2nd ed. Prentice Hall, Englewood Cliffs, NJ. ISBN 9780137095926.
Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G.A., Zhu, Z., Gong, W., 2016. Solar radiation prediction using different techniques: model evaluation and comparison. Renew. Sust. Energ. Rev. 61, 384–397.
Wang, L., Kisi, O., Zounemat-Kermani, M., Zhu, Z., Gong, W., Niu, Z., Liu, Z., 2017. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. Int. J. Climatol. 37 (3), 1141–1155.
Yadav, A.K., Malik, H., Chandel, S.S., 2014. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew. Sust. Energ. Rev. 31, 509–519.
Yıldırım, H.B., Çelik, O., Teke, A., Barutçu, B., 2018. Estimating daily global solar radiation with graphical user interface in Eastern Mediterranean region of Turkey. Renew. Sust. Energ. Rev. 82, 1528–1537.
Yi-zhou, Lin, Zhen-hua, Nie, Hong-wei, Ma, 2017. Structural damage detection with automatic feature-extraction through deep learning. Comput Aided Civ Infrastruct Eng 32 (12), 1025–1046.
Yuming, Hua, Junhai, Guo, Hua, Zhao, 2015. Deep belief networks and deep learning. In: Proceedings of 2015 international Conference on Intelligent Computing and Internet of Things. IEEE.
Zang, H., Xu, Q., Bian, H., 2012. Generation of typical solar radiation data for different climates of China. Energy 38 (1), 236–248.
Zeng, Z., Wang, Z., Gui, K., Yan, X., Gao, M., Luo, M., et al., 2020. Daily global solar radiation in China estimated from high-density meteorological observations: a random forest model framework. Earth Space Sci. 7 https://doi.org/10.1029/2019EA001058 e2019EA001058.