Renewable and Sustainable Energy Reviews 135 (2021) 110114

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: http://www.elsevier.com/locate/rser

# Prediction of daily global solar radiation using different machine learning algorithms: Evaluation and comparison

Ümit Ağbulut [a], Ali Etem Gürel [a,b,*], Yunus Biçen [c]

[a] Department of Mechanical Engineering, Faculty of Technology, Düzce University, 81620, Turkey
[b] Department of Electricity and Energy, Vocational School, Düzce University, 81010, Turkey
[c] Department of Electrical and Electronics Engineering, Faculty of Technology, Düzce University, 81620, Turkey

ARTICLE INFO

*Keywords:*
Daily global solar radiation
Machine learning algorithms
Solar energy
Prediction
Model comparison

ABSTRACT

The prediction of global solar radiation for the regions is of great importance in terms of giving directions of solar energy conversion systems (design, modeling, and operation), selection of proper regions, and even future investment policies of the decision-makers. With this viewpoint, the objective of this paper is to predict daily global solar radiation data of four provinces (Kırklareli, Tokat, Nevşehir and Karaman) which have different solar radiation distribution in Turkey. In the study, four different machine learning algorithms (support vector machine (SVM), artificial neural network (ANN), kernel and nearest-neighbor (k-NN), and deep learning (DL)) are used. In the training of these algorithms, daily minimum and maximum ambient temperature, cloud cover, daily extra-terrestrial solar radiation, day length and solar radiation of these provinces are used. The data is supplied from the Turkish State Meteorological Service and cover the last two years (01.01.2018–31.12.2019). To decide on the success of these algorithms, seven different statistical metrics ($R^2$, RMSE, rRMSE, MBE, MABE, t-stat, and MAPE) are discussed in the study. The results shows that $R^2$, MABE, and RMSE values of all algorithms are ranging from 0.855 to 0.936, from 1.870 to 2.328 MJ/m$^2$, from 2.273 to 2.820 MJ/m$^2$, respectively. At all cases, k-NN exhibited the worst result in terms of $R^2$, RMSE, and MABE metrics. Of all the models, DL was the only model that exceeded the t-critic value. In conclusion, the present paper is reporting that all machine learning algorithms tested in this study can be used in the prediction of daily global solar radiation data with a high accuracy; however, the ANN algorithm is the best fitting algorithm among all algorithms. Then it is followed by DL, SVM and k-NN, respectively.

## 1. Introduction

The sun is an abundant, free, infinite, green, reliable, renewable, and sustainable energy resource. It also has a large potential to fulfill the energy demands of the world and it is projected that it will be widely utilized in the immediate future [1–3]. In order to determine the solar energy potential of a region, the solar radiation (SR) information of that region must be known [4,5]. For robust planning, management and applications of investments in the solar energy field, a successful prediction of solar radiation amount coming onto the ground is one of the remarkable issues [6,7]. In the case that the energy produced after the investments is sold commercially, it is important to make accurate particularly short-term solar radiation prediction both in the day ahead and intraday markets. Making accurate predictions directly affects the profit margin of the energy suppliers in these markets [8–10]. The term "solar radiation" throughout this study refers to "horizontal surface global solar radiation" if not otherwise stated.

In general, solar radiation is measured by the aid of SR measurement instruments. However, these instruments have high installation-maintenance costs and calibration requirements. Therefore, they are not available at most stations worldwide. For instance, until 2020, there are 1798 meteorological stations in Turkey, and only 129 of them is able to record the solar radiation data [11]. Similarly, until 2012, there are 756 stations in China and only 122 of them is able to record the solar radiation data [12]. That is why it is of significantly importance to predict the solar radiation data with the easily measured climatic parameters such as; humidity, temperature, wind speed, cloud cover, etc. With this viewpoint, a numerous model has been proposed to predict the solar radiation data. Some of them are based on mathematical formulas and called empirical models. Empirical models are easy to calculate and accepted as a useful technique to predict solar radiation data. Even

* Corresponding author. Department of Mechanical Engineering, Faculty of Technology, Düzce University, 81620, Turkey.
*E-mail addresses:* umitagbulut@duzce.edu.tr (Ü. Ağbulut), alietemgurel@duzce.edu.tr (A.E. Gürel), yunusbicen@duzce.edu.tr (Y. Biçen).

https://doi.org/10.1016/j.rser.2020.110114
Received 15 March 2020; Received in revised form 2 July 2020; Accepted 17 July 2020
Available online 18 August 2020
1364-0321/

**Nomenclature**

| | |
|---|---|
| AI | Artificial intelligence |
| ANFIS | Adaptive neuro-fuzzy inference system |
| ANN | Artificial neural network |
| ARMA | Autoregressive and moving average |
| BC | Bristow–Campbel |
| φ | Latitude angle of the region |
| D | Number of days |
| DL | Deep learning |
| DT | Desicion tree |
| GEP | Gene expression programming |
| GPR | Gaussian process regression |
| GRNN | Generalized regression neural network |
| $I_{sc}$ | Solar constant |
| KELM | Kernel extreme learning machine |
| k-NN | Kernel nearest neighbor |
| MABE | Mean absolute bias error |
| MAPE | Mean absolute percentage error |
| MBE | Mean bias error |
| MEA | Mind evolutionary algorithm |
| MEA-ANN | Hybrid mind evolutionary algorithm and artificial neural network |
| MENR | Turkey ministry of energy and natural resources |
| ML | Machine learning |
| MLP | Multi-layer perceptron |
| MLR | Multi-linear regression |
| n | Number of observations |
| NA | Not available |
| $R^2$ | Coefficient of determination |
| RBF | Radial basis function |
| RF | Random fores |
| RMSE | Root mean squared error |
| rRMSE | Relative root mean squared error |
| $S_o$ | Day length |
| SR | Solar radiation |
| SVM | Support vector machine |
| SVR | Support vector regression |
| $T_{max}$ | Maximum ambient temperature |
| $T_{min}$ | Minimum ambient temperature |
| TSMS | Turkish state meteorological service |
| t-stat | t statistic |
| WNN | Wavelet neural network |
| WT | Wavelet Transform |
| δ | Declination angle |
| $\bar{x}_i$ | Mean of measured daily global solar radiation |
| $x_i$ | Measured daily global solar radiation |
| $y_i$ | Predicted daily global solar radiation |
| $\omega_s$ | Sunset hour angle |

though empirical models have been frequently used in the prediction of monthly average daily global solar radiation, these models are not capable of accurately predicting the short-term solar radiation data because of the rapid changes in weather conditions such as; cloud cover, rainy days, etc. In line with this, some researchers reported that these models are not also capable of reflecting the complex and nonlinear relationships among both dependent and independent variables in humid regions in which solar radiation is strongly affected by heavy clouds throughout rainy days [13,14]. It was reported in the previous studies that these empirical models have presented partially-unsatisfying prediction results for the daily global solar radiation data [15–19].

With technological improvements, in recent decades, artificial intelligence (AI) has become very popular among almost all engineering fields [20,21]. In addition to empirical models, different AI methods such as support vector machine (SVM), deep learning (DL), kernel nearest neighbor (k-NN), artificial neural networks (ANN), genetic algorithms (GA), etc. started to be frequently used models in the prediction of solar radiation data. The previous studies are reporting that the AI algorithms have presented more accurate results than those of the empirical models for the prediction of solar radiation [13,22,23]. For example, Quej et al. predicted daily global solar radiation data of six stations in Mexico by using three machine learning algorithms, namely SVM, ANN, and ANFIS. The authors used extraterrestrial solar radiation, rainfall, minimum temperature, and temperature data for the training of the algorithms. In the relevant study, the best results were achieved in SVM with RMSE = 2.578, MAE = 1.97 and $R^2 = 0.689$ [24]. In another study, Marzo et al. tried to predict the daily global solar radiation of 13 different stations. The authors only used ANN as a machine learning algorithm in that study. In the study, extraterrestrial solar radiation, minimum temperature, and maximum temperature were used to train the ANN algorithm. In the results, rRMSE = 13%, rMBE<4% and r = 0.800 ($R^2 = 0.64$) were calculated as the best results [25]. Mehdizadeh et al. used three different models, namely Gene Expression Programming (GEP), ANN and ANFIS, for the daily global solar radiation in Karmen, Iran. The best results in the relevant study were seen in the ANN model. In this model, $R^2$ was calculated to be 0.935 [26]. Tymvios et al. compared to Angström model and ANN method for the prediction of global solar radiation data. In the results, the authors found that the ANN method presented better prediction results in comparison with Angström model [27]. In another study, Meenal and Selvakumar studied the daily global solar radiation with SVM, empirical and ANN models. Among these models, SVM presented the best results with more than 0.99 correlation [89]. Yildirim et al. studied on the prediction of the monthly global solar radiation with regression analysis and machine learning algorithm (ANN) for four different stations in Turkey. In the relevant study, extraterrestrial solar radiation, longitude, maximum possible sunshine duration, sunshine duration, relative humidity, temperature and day of the year were used for the training the algorithm. The authors have achieved the best results in the ANN model with 0.961 and 0.14 of $R^2$ and RMSE, respectively [4]. Kaba et al. studied the prediction of monthly average daily global solar radiation of different stations in Turkey with an only deep learning algorithm. In the study, extraterrestrial solar radiation, sunshine duration, cloud cover, minimum and maximum temperature data were used to train the DL algorithm. In the relevant study, $R^2$ value was equal to 0.98 in the best results [28]. In another study, Marzouq et al. used k-NN, ANN and empirical models for the prediction of daily global solar radiation. In the results, $R^2$ value of k-NN was equal to 0.96. On the other hand, a proposed hybrid model (k-NN and ANN) had the 0.97 of $R^2$ in the relevant paper [29].

As mentioned earlier, even if there are many meteorological stations in the countries, the vast majority of them are not able to measure solar radiation data due to the cost of solar radiation measurement equipment, maintenance, and calibration thereof [10]. It is, therefore, significant to predict the solar radiation data by using some variables being easier to measure in the relevant region. As can be seen from previous literature works, different machine learning algorithms are frequently used in the prediction of daily global solar radiation recently. The main reason why these algorithms are frequently used is that more accurate prediction results, particularly for daily global solar radiation, could be achieved with these algorithms in comparison with those of empirical models. However, it is also seen from the previous studies that the algorithms tested are generally giving the close results with each other for

**Table 1**
Some important geographical details of the selected provinces.

| Location | Latitude | Longitude | Elevation (m) |
|---|---|---|---|
| Kırklareli | 41° 44′ N | 27° 13′ E | 231 |
| Nevşehir | 38° 36′ N | 34° 42′ E | 1197 |
| Tokat | 40° 19′ N | 36° 33′ E | 630 |
| Karaman | 37° 11′ N | 33° 13′ E | 1063 |

the study sites. Therefore, using a few metrics to discuss the performance success of the algoriths may be insufficient. With these viewpoints, this research contributes to the literature in the following three ways:

- **Comparison of frequently and infrequently used four machine learning algorithms on the same dataset:** In addition to the frequently used ANN and SVM algorithms, DL and k-NN algorithms are also used in the prediction of daily global solar radiation data in this study.
- **Prediction of Turkey's overall solar radiation distribution:** Four different provinces with very-low, low, medium and high solar radiation potential are used in this paper. These provinces are representing Turkey's overall solar radiation distribution.
- **A comprehensively discussion of the results:** Unlike many studies comparing the prediction successes with several metrics, this paper presents a comprehensive discussion on the algorithm success using seven metrics together ($R^2$, RMSE, rRMSE, MBE, MABE, MAPE, and t-stat).

Accordingly, the remaining of this paper is organized as follows: Section 2 gives the details about study regions, data collection, intoruction of the machine learning algorithms, and statistical metrics. Section 3 presents the results of the metrics and prediction graphs. Also, metric comparison of present paper with the literature studies in the prediction of daily global solar radiation are compheresevely evaulated in Section 3. Finally, Section 4 discusses the conslusion.

## 2. Material and method

This section is divided into four subsections. Firstly, the information about the study sites is given. In the second stage, information about the data collection and pre-processing is mentioned. A brief summarization regarding the machine learning algorithms is given. Finally, evaluation metrics are presented in the last subsection.

### 2.1. Study sites

Turkey is a country located between the continents of Europe, Africa, and Asia. Its longitudes and latitudes are 26–45 and 36–42, respectively. Turkey has 81 provinces and its total area is 783562 km$^2$. Turkey has a vast solar energy potential as compared to European countries. According to the report of the Republic of Turkey Ministry of Energy and Natural Resources (MENR), annual global solar radiation and annual sunshine duration of Turkey are 1527 kWh/m$^2$-year and 2471 h and, respectively. This case puts Turkey to an attractive location for solar

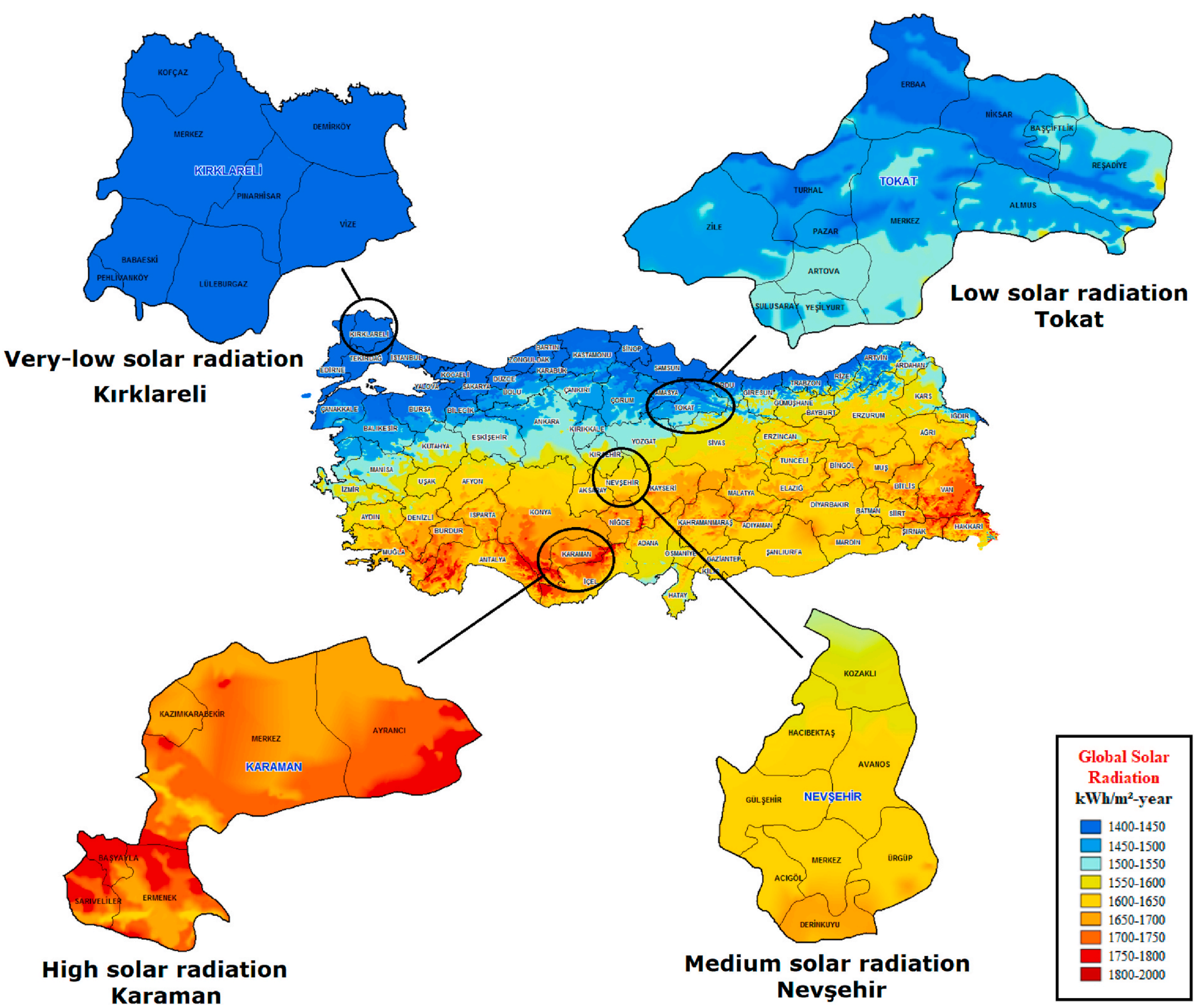


**Fig. 1.** Annual global solar radiation distribution of the provinces on Turkish map [30].

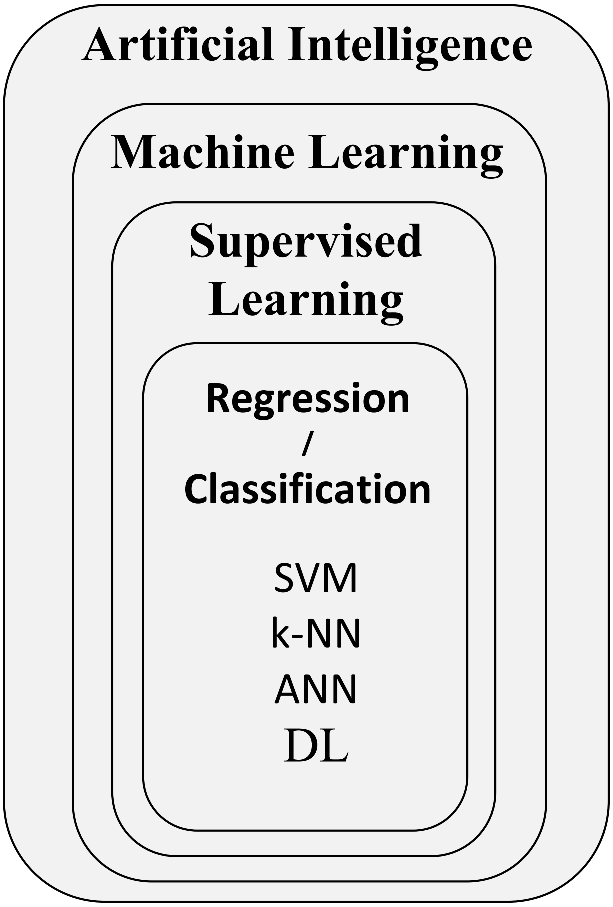


**Fig. 2.** Core structure of the study.

energy investments.

In this paper, four provinces in Turkey were selected for the prediction of daily global solar radiation. Some important geographical details and the view of the provinces on the Turkish map are given in Table 1 and Fig. 1, respectively.

As can be seen from Fig. 1, global solar radiation of the four provinces used in this study has different scales, that is, Kirklareli province has almost at least global solar radiation among all provinces in Turkey. On the other hand, Tokat province can be considered as a transition region in terms of global solar radiation and the last province – Nevşehir – has partially-higher global solar radiation in comparison with the other two provinces. Finally, Karaman is the province with the highest global solar radiation. With this viewpoint, these four provinces adopted in the present paper can represent Turkey's overall solar radiation.

### 2.2. Data collection

The present paper focuses on the prediction of daily average global solar radiation data for four different provinces in Turkey. In the prediction, minimum and maximum ambient temperature ($T_{min}$ and $T_{max}$), cloud cover and solar radiation, which were daily measured from four stations for the last two years, were included in the study. In addition to these, daily extraterrestrial solar radiation and day length were calculated by using Eqs. (1)–(4) [31]. This dataset was supplied from the Turkish State Meteorological Service [11] and it covers the term from January 01, 2018 to December 31, 2019.

$$H_o = \frac{24}{\pi} I_{sc}\left(1 + 0.033cos\frac{360D}{365}\right) \times \left(cos(\varphi)cos(\delta)sin(\omega_s)\frac{2\pi\omega_s}{360}sin(\varphi)sin(\delta)\right) \quad (1)$$

where $I_{sc}$ solar constant (1367 W/m$^2$), φ is latitude angle of the region, δ is the declination angle, $\omega_s$ is the sunset hour angle and $D$ represents the number of days from 1 January.

The declination angle and sunset hour angle can be calculated using the following equations [31].

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

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

The day length to be calculated can be stated as follows [31]:

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

### 2.3. Machine learning algorithms

Machine learning (ML) is a frequently used type of artificial intelligence (AI) and it continues to be popular and attractive by finding various application fields day by day. ML provides the systems the ability to understand by itself and then to estimate the unknown outputs. Undoubtedly, the performance of an ML algorithm is strongly dependent on the selection of attributes and training success of it. In this study, four different ML algorithms were used. As shown in Fig. 2, these are Artificial Neural Network (ANN), Kernel and Nearest Neighbor (k-NN), Support Vector Machine (SVM), and Deep Learning (DL).

All the algorithms were performed by using RapidMiner Studio Version 9.5. In addition, as mentioned before, five different attributes regarding the daily global solar radiation were collected from the meteorological stations or calculated. These are the daily minimum and maximum ambient temperature ($T_{min}$ and $T_{max}$), cloud cover, daily extraterrestrial solar radiation and day length. In the algorithms, the dataset was randomly split as the shuffled sampling mode, and 70% of the total data was used in the training stage of the algorithms while the remaining 30% was used in the testing stage. In all methods, the same dataset for each station was used both for the training and testing data, and the same labels were predicted in order to provide a better comparison among the ML models.

#### 2.3.1. Artificial neural network (ANN)

ANN is an ideal solution tool for modeling systems with non-linear relationship and it is a widely used method both for regression and classification problems. So there is no need for complex mathematical expressions to describe such systems [32,33]. ANN is a network structure consisting of elements with process capability called neurons. However, the created neural network structure must be trained to establish the relationship between inputs and outputs [32,34]. In this respect, it is very similar to the information processing mechanism of the human brain [35]. After training, they can produce fast and reliable solutions even in datasets containing noise and missing information. If the complexity of the systems is high, the data set required for training may be more than needed in simple methods and some preprocessing may be required in the data set [36,37]. Depending on the size of the network, the number of hidden layers can also grow in ANN. This can cause the network to memorize the data set and even noise [32,38,39]. It is important to be able to set the optimum network size to avoid this situation and reduce training time. In this paper, the ANN used a feed-forward neural network and trained by means of the back-propagation algorithm (multi-layer perceptron) for all provinces. In this research, the best results for ANN are observed when the hidden layer is 14 for Kırklareli, 13 for Tokat, 8 for Nevşehir, and 16 for Karaman.

#### 2.3.2. Kernel nearest neighbor (k-NN)

k-NN is known as one of the simplest and oldest non-parametric supervised classification approaches among machine learning algorithms in the literature [40,41]. By defining a special number k in the total data set, the average/modes classes of the nearest neighbors are obtained and the new object is assigned to the nearest class to its neighbors [42,43]. The distances of the new object with its neighbors can be calculated with functions such as Euclidean, Manhattan, Minkowski, and Chebyshev [44]. It has a robust structure against noisy training data provided that the number k is large enough [45,46]. When the data set and k size increase, the processing time to be performed increases considerably and also in this approach it has to be kept all these distance calculations results in memory [47,48]. Therefore, the choice of the k value is extremely important. In this research, the best prediction result is observed in which k is 7, 12, 11, and 8 for Kırklareli, Nevşehir, Tokat, and Karaman, respectively.

#### 2.3.3. Support vector machine (SVM)

While support vector machine was originally used to separate the two classes, it was developed over time and successfully used in regression, classification and outlier detection problems with nonlinear systems [49–51]. It is a supervised parametric machine learning algorithm based on statistical learning theory [52,53]. In order to separate the two classes in the SVM algorithm, a parallel line/hyperplane is drawn between the data that make up the classes [54]. The structure used to separate classes is shown as a line in two-dimensional space, while in three-dimensional space it is shown as a plane [55]. The data closest to the hyperplane are called support vectors. The margin between the support vectors of opposite classes is maximized, thus it becomes more noise resistant [55,56]. In this paper, radial was assigned as kernel type in SVM algorithm for all provinces. In this research, kernel cache, c-epsilon, and maximum iteratiron values are 300, 1. E6, and 10000000, respectively for all algorithms. However, kernel gamma and C parameters changed to be 0.05 and 0, respectivey for Kırklareli, 0.15 and 35, respectivey for Tokat, 0.15 and 20, respectivey for Nevşehir, 0.15 and 35, respectivey for Tokat, 0.25 and 200, respectivey for Karaman.

#### 2.3.4. Deep learning (DL)

DL is a popular tool used in the solution of complex problems that contain particularly large data sets [57,58]. In deep learning, supervised, semi-supervised and unsupervised operations can be performed [59–63]. Unlike classical machine learning approaches, DL has the ability to perform feature extraction itself, even if raw data is given as input and in this approach, the increasing of the data set means the rising of learning performance [64,65]. DL's outstanding success owes its neural network structure consisting of many hidden layers [66,67]. Depending on the number of layers and the size of the data set, the learning time is longer compared to other machine learning algorithms [68,69]. In this paper, the rectifier was taken as the activation of DL. In the research, epsilon is taken 1. E9 and loss function are chosen as quandratic for all algorithms. The best results are observed in which epoch and rho parameters are 10 and 0.99, respectively for Kırklareli, 11 and 0.9, respectively for Nevşehir, 11 and 0.9, respectively for Tokat, 0.15 and 15, respectivey for Karaman.

### 2.4. Evaluation metrics

Accuracy is clearly the most significant criterion in the evaluation of the performance success of the prediction methods. Therefore, the commonly used error metrics are used both in evaluating the results of the prediction models and in comparison with each other. Some metrics such as MBE (Mean bias error), RMSE (Root mean squared error), rRMSE (Relative root mean squared error), t-stat, MABE (Mean absolute bias error), $R^2$ (Coefficient of determination), and MAPE (Mean absolute percentage error) were used to compare the performance success of the prediction models used in the present paper.

These statistical metrics, their equations, and descriptions are given in Table 2.

In Table 2, $y_i$and $x_i$are predicted and measured daily global solar radiation, respectively; $\overline{x}_i$is the mean of measured daily global solar radiation; n is the number of observations.

**Table 2**
A brief summary of the statistical metrics used in the study.

| Metrics | Equation | Description |
|---|---|---|
| MBE | $\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)$ | MBE is an important metric for the long-term performance of prediction models. The small value of MBE indicates that the prediction model has better performance. Also, zero represents the ideal case [13]. |
| RMSE | $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}$ | RMSE provides information on the short-term performance of the prediction models. Its value is always positive and is desired to be close to zero [70]. |
| rRMSE | $\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i)^2}}{\overline{x}_i} \times 100$ | rRMSE is obtained by the RMSE and the average value of the measured data. The small value of rRMSE indicates that the used prediction model has better performance [40]. The success of the prediction model was considered as follows:<br>Excellent: rRMSE<10%<br>Good: 10%<rRMSE<20%<br>Fair: 20%<rRMSE<30%<br>Poor: rRMSE>30% [71,72,87]. |
| t-stat | $\sqrt{\frac{(n-1)MBE^2}{RMSE^2 - MBE^2}}$ | t-stat is applied to indicate if the predictions of a model are statistically meaningful. The smaller the value of t-stat, the better is the prediction model performance. In this method, a t-critic value from statistical tables is determined [31,87]. |
| MABE | $\frac{1}{n}\sum_{i=1}^{n}\lvert y_i - x_i\rvert$ | MABE is the absolute value of the bias error and is a measure of the goodness of a correlation. Its desired to be close to zero. MABE provides knowledge about the long-term performance of the prediction models [73,74]. |
| $R^2$ | $1 - \frac{\sum(x_i - y_i)^2}{\sum(x_i - \overline{x}_i)^2}$ | This approach provides knowledge about how well a model can predict a set of measured data. Its value varies between 0 and 1. The $R^2$ value approaching 1 is an indication of better performance [75]. |
| MAPE | $\frac{1}{n}\sum_{i=1}^{n}\left\lvert\frac{x_i - y_i}{x_i}\right\rvert \times 100$ | MAPE is the percentage of the average of the absolute values of prediction errors to the absolute values of actual data. The lower value of the MAPE is an indicator of better performance of the model [70,76]. The success of the prediction model as follows:<br>High prediction accuracy: MAPE≤10%<br>Good prediction: 10%<MAPE≤20%<br>Reasonable prediction: 20%<MAPE≤50%<br>Inaccurate prediction: MAPE>50% [77]. |

## 3. Results and discussions

The present paper deals with the predictability of four different provinces in Turkey of daily solar radiation falling onto the horizontal surface via four different machine learning algorithms. In order to evaluate the success of these algorithms, seven different statistical metrics, which are frequently used in the literature, are discussed, Table 3 gives the numerical values of the metrics calculated both for all provinces and algorithms in the study. As seen in Table 3, the coefficient of determination ($R^2$) is varying between 0.855 and 0.932 depending on the province and algorithm. In other words, it can be said that all algorithms in terms of $R^2$ exhibit good performance in the prediction of the daily global solar radiation. In this section, all provinces and

**Table 3**
Performance comparison of provinces for each algorithm.

| Provinces | Metric | SVM | k-NN | DL | ANN |
|---|---|---|---|---|---|
| **Kırklareli** | $R^2$ | 0.898 | 0.890 | 0.915 | 0.913 |
| | RMSE ($MJ/m^2$) | 2.651 | 2.748 | 2.417 | 2.443 |
| | rRMSE (%) | 18.50 | 19.17 | 16.86 | 17.05 |
| | MBE ($MJ/m^2$) | −0.039 | 0.038 | 0.180 | 0.231 |
| | MABE ($MJ/m^2$) | 1.698 | 1.870 | 1.505 | 1.547 |
| | t-stat | 0.207 | 0.194 | 1.050 | 1.333 |
| | MAPE (%) | 19.79 | 26.18 | 23.71 | 22.56 |
| **Tokat** | $R^2$ | 0.880 | 0.855 | 0.875 | 0.883 |
| | RMSE ($MJ/m^2$) | 2.820 | 3.092 | 2.814 | 2.776 |
| | rRMSE (%) | 18.43 | 20.21 | 18.39 | 18.14 |
| | MBE ($MJ/m^2$) | 0.289 | 0.209 | 0.659 | 0.138 |
| | MABE ($MJ/m^2$) | 1.968 | 2.328 | 2.075 | 2.054 |
| | t-stat | 1.439 | 0.945 | 3.288 | 0.698 |
| | MAPE (%) | 23.37 | 27.74 | 26.61 | 23.33 |
| **Nevşehir** | $R^2$ | 0.926 | 0.902 | 0.932 | 0.932 |
| | RMSE ($MJ/m^2$) | 2.273 | 2.531 | 2.203 | 2.157 |
| | rRMSE (%) | 14.85 | 16.54 | 14.40 | 14.10 |
| | MBE ($MJ/m^2$) | −0.012 | 0.093 | 0.676 | 0.195 |
| | MABE ($MJ/m^2$) | 1.736 | 1.953 | 1.712 | 1.597 |
| | t-stat | 0.075 | 0.518 | 4.341 | 1.280 |
| | MAPE (%) | 16.03 | 22.06 | 22.25 | 15.92 |
| **Karaman** | $R^2$ | 0.926 | 0.886 | 0.928 | 0.936 |
| | RMSE ($MJ/m^2$) | 2.414 | 2.992 | 2.390 | 2.218 |
| | rRMSE (%) | 20.32 | 25.19 | 20.11 | 18.67 |
| | MBE ($MJ/m^2$) | −0.0678 | 0.2272 | 0.0346 | −0.0251 |
| | MABE ($MJ/m^2$) | 1.8124 | 2.197 | 1.822 | 1.912 |
| | t-stat | 0.3983 | 1.068 | 0.2054 | 0.1608 |
| | MAPE (%) | 20.32 | 30.24 | 24.18 | 17.82 |

algorithms will be compared and discussed by considering Table 3 as the reference.

Fig. 3 a shows the daily solar radiation data and error magnitude both measured for the Kırklareli region and predicted by the SVM algorithm. Considering Fig. 3a and Table 3 together, it is seen that the $R^2$ value of SVM algorithm is equal to 0.898. Among all algorithms, SVM algorithm gives the most successful prediction results in terms of MAPE for Kırklareli province. In addition, the predictive results of this algorithm are meaningful in terms of t-stat (for t-critical value: 2.60, $n = 206$, $\alpha = 0.01$). The SVM algorithm is the single-algorithm that presented negative results for Kırklareli province in terms of MBE. In other words, the average of SR predictions by SVM is lower than the average of actual SR values. Three observations with the most error magnitudes for SVM were 55 (error = 19.62 $MJ/m^2$), 136 (error = 12.30 $MJ/m^2$) and 68 (error = 6.93 $MJ/m^2$), respectively.

Fig. 3 b shows the daily solar radiation data and error magnitude both measured for the Kırklareli region and predicted by the k-NN algorithm. The k-NN algorithm has been the most unsuccessful algorithm in terms of statistical metrics in predicting solar radiation data for Kırklareli province. Even if the algorithm with the worst $R^2$, RMSE and MAPE values is k-NN, this algorithm is also meaningful in terms of t-stat. Similar to the SVM algorithm, the k-NN algorithm had the biggest error magnitude in observation 55 (error = 19.69 $MJ/m^2$). Unlike SVM, in the predictions by the k-NN algorithm, a considerable error was seen in observation number 77 (error = 6.75 $MJ/m^2$).

Fig. 3 c shows the daily solar radiation data and error magnitude both measured for the Kırklareli region and predicted by the DL algorithm. Among all the algorithms for Kırklareli province, the most successful algorithm in terms of $R^2$, RMSE and MABE metrics is DL. The t-stat value for this algorithm was calculated to be higher both than SVM and k-NN algorithms but it is also meaningful in terms of t-stat. The reason why the t-stat value calculated higher than SVM and k-NN algorithms is due to the higher MBE values calculated for DL compared to the SVM and k-NN algorithms. Considering the error magnitudes of the DL algorithm, unlike SVM and k-NN algorithms, it is seen that the error magnitude is less in observations from 1 to 54. Likewise, in recent observations (particularly observations of 181 and further), the error magnitudes of the DL algorithm were seen to be quite low.

Fig. 3 d shows the daily solar radiation data and error magnitude both measured for the Kırklareli region and predicted by the ANN algorithm. To give a general assessment for Kırklareli province, DL and ANN algorithms showed a very close prediction success. Particularly in terms of $R^2$, RMSE, MABE, and MAPE metrics, DL and ANN are very close to each other. The error magnitudes of both algorithms are parallel to each other. Considering the success of all algorithms in terms of rRMSE, it is seen that the results of all algorithms can be categorized as “good”, and the most successful metric in terms of rRMSE is DL with 16.86%.

Fig. 4 a shows the daily solar radiation data and error magnitude both measured for the Tokat region and predicted by the SVM algorithm. Even if it gives satisfying results in terms of statistical metrics, Tokat is the province where all algorithms performed the worst prediction results among four provinces selected in this study. This situation can be clearly seen from the error magnitudes of the prediction. Many observations have high error magnitudes and this has increased the total error for this province. The reason why Tokat is a less predictable province among these four provinces may probably be due to the fact that the solar radiation data are highly changeable in comparison with other provinces and so this province was called a transition zone above in terms of solar radiation (see Fig. 1). SVM algorithm has been the most successful prediction algorithm with ANN in terms of statistical metrics.

Fig. 4 b shows the daily solar radiation data and error magnitude both measured for the Tokat region and predicted by the k-NN algorithm. The k-NN algorithm exhibits the most unsuccessful performance in terms of $R^2$, RMSE, MABE and rRMSE statistical metrics for Tokat province. The k-NN algorithm has higher error magnitudes, particularly in the observation ranges 79–137 and 185–206 in comparison with that of SVM. These errors also negatively affected the success of this algorithm.

Fig. 4 c shows the daily solar radiation data and error magnitude both measured for the Tokat region and predicted by the DL algorithm. As seen in Table 3, the DL algorithm, in terms of $R^2$, was the second most unsuccessful algorithm for Tokat province. DL was the worst algorithm in terms of both MBE of 0.659 $MJ/m^2$ value and t-stat. The MBE value, which is quite high compared to the other three algorithms, caused the DL algorithm to exceed the t-critical value.

Fig. 4 d shows the daily solar radiation data and error magnitude both measured for the Tokat region and predicted by the ANN algorithm. For Tokat province, ANN has been the algorithm that gives the most successful prediction in terms of all statistical metrics. As seen in Table 3, the ANN algorithm has the highest $R^2$ value (0.883) and the lowest RMSE value (2.776 $MJ/m^2$) for this province. The most smaller rRMSE value is obtained with 18.14% in the ANN algorithm. Considering the error magnitudes of SVM and ANN together, which are the two best-performing algorithms for Tokat province, it is seen that ANN has a very low error magnitude particularly in observations between 1 and 79 and therefore ANN comes to the fore from SVM.

Fig. 5 a shows the daily solar radiation data and error magnitude both measured for the Nevşehir region and predicted by the SVM algorithm. Considering four provinces within the scope of the study, Nevşehir is the province where the algorithms showed the highest performance in terms of many statistical metrics. The SVM algorithm has shown quite satisfactory results for this province. This algorithm for Nevşehir province had $R^2$ value of 0.926, it also had the lowest MBE value with −0.012 $MJ/m^2$. On the other hand, SVM is the single algorithm with a negative MBE value for Nevşehir province.

Fig. 5 b shows the daily solar radiation data and error magnitude both measured for the Nevşehir region and predicted by the k-NN algorithm. Even if the k-NN algorithm for Nevşehir has a partially-high $R^2$ value (0.902), it has the worst prediction performance in terms of other metrics. For example, as it can be seen in Table 3, k-NN has the highest RMSE value (2.531 $MJ/m^2$).

Fig. 5c and 5 d shows the daily solar radiation data and error

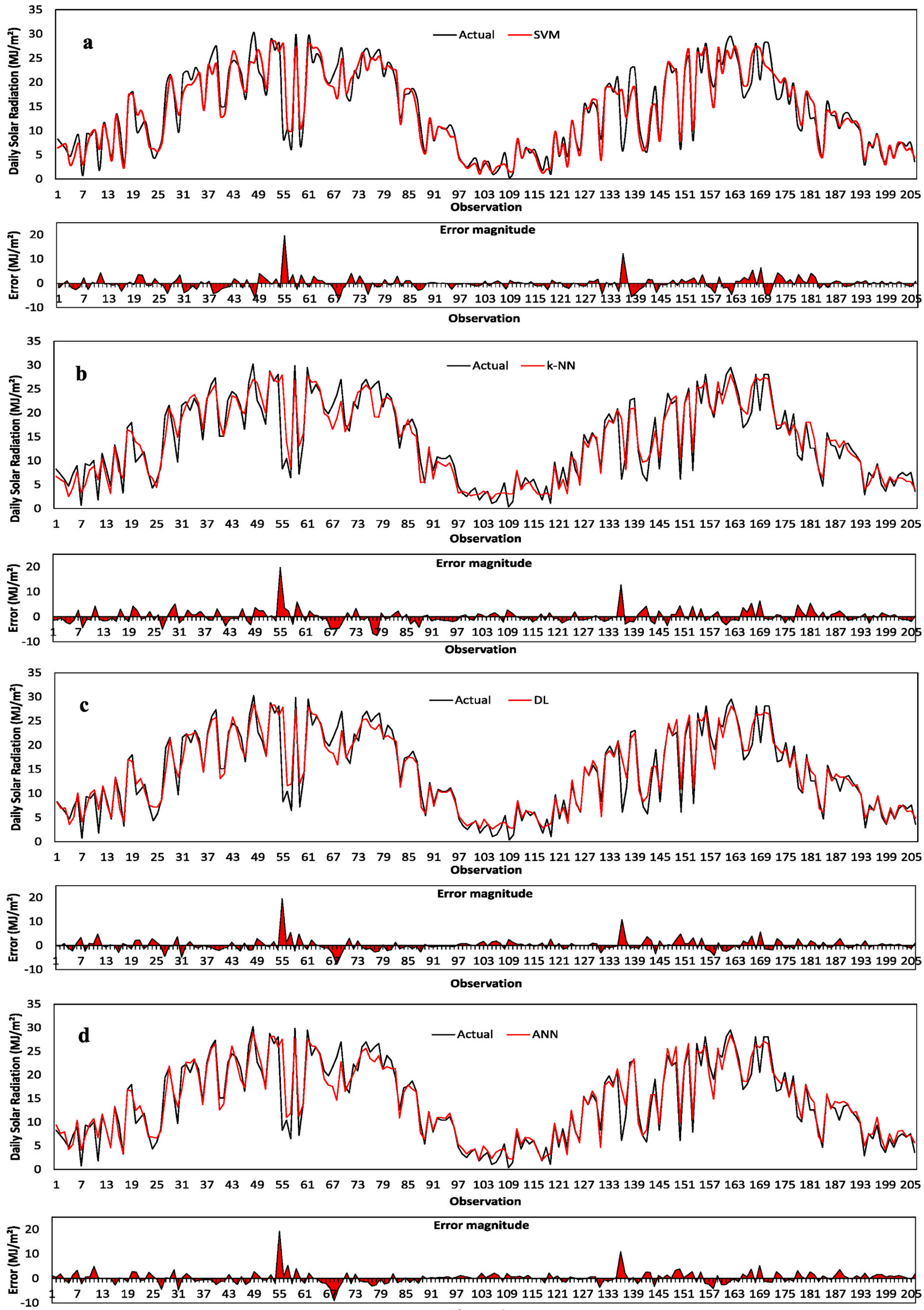


**Fig. 3.** Actual SR data, prediction result and error magnitude for Kırklareli province a) SVM b) k-NN c) DL, and d) ANN (testing data of 30% for each year).

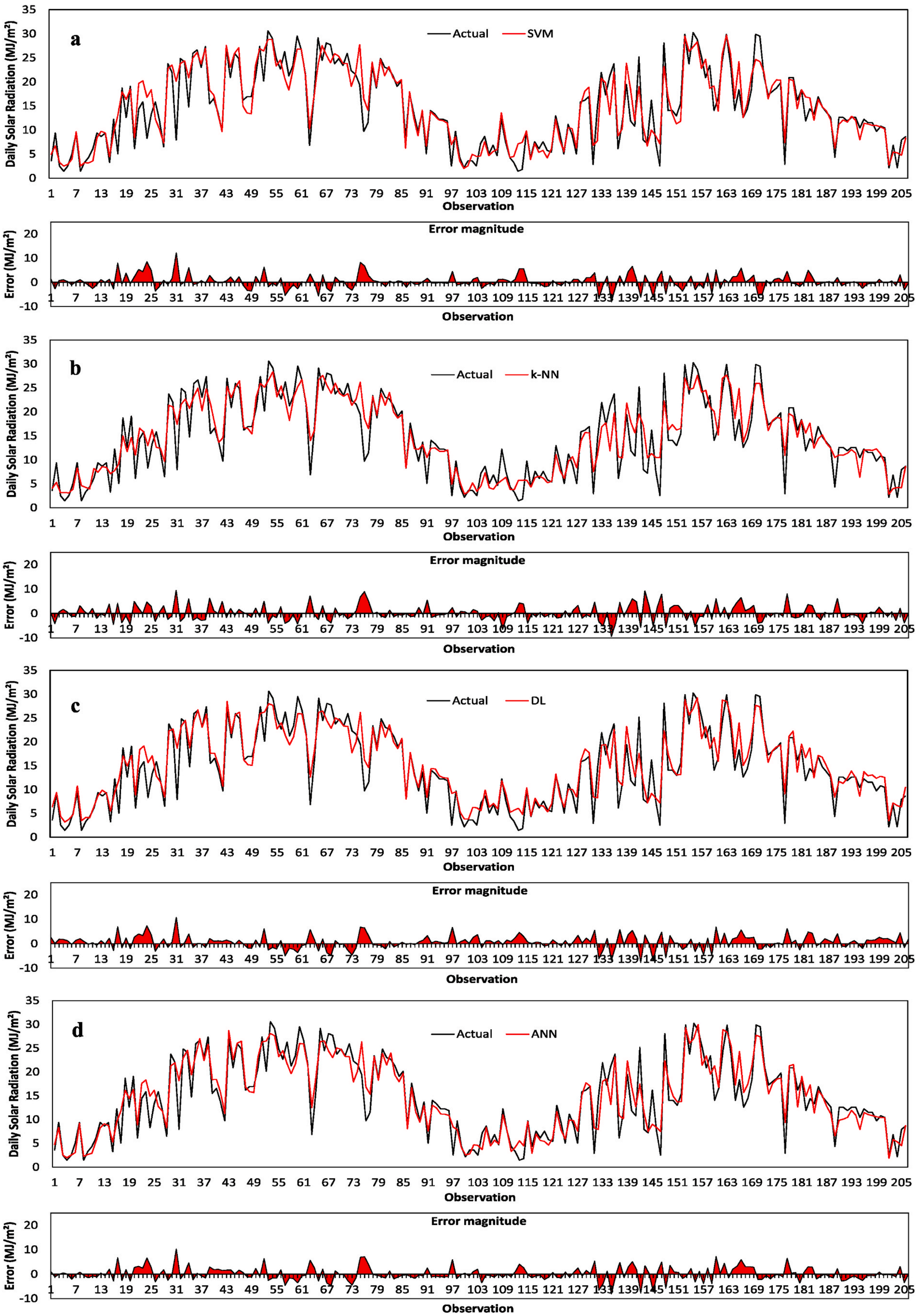


**Fig. 4.** Actual SR data, prediction result and error magnitude for Tokat province a) SVM b) k-NN c) DL, and d) ANN (testing data of 30% for each year).

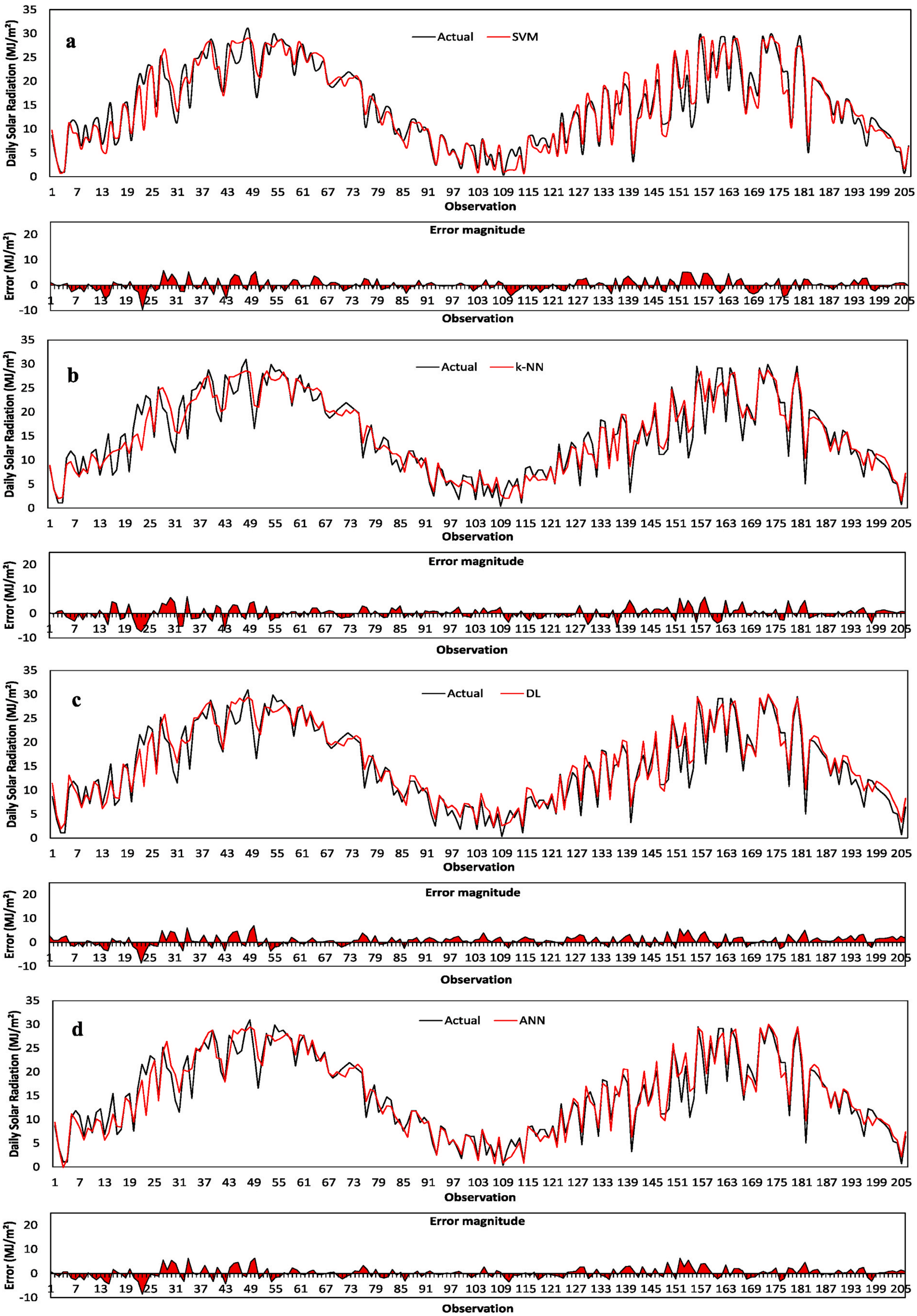


**Fig. 5.** Actual SR data, prediction result and error magnitude for Nevşehir province a) SVM b) k-NN c) DL, and d) ANN (testing data of 30% for each year).

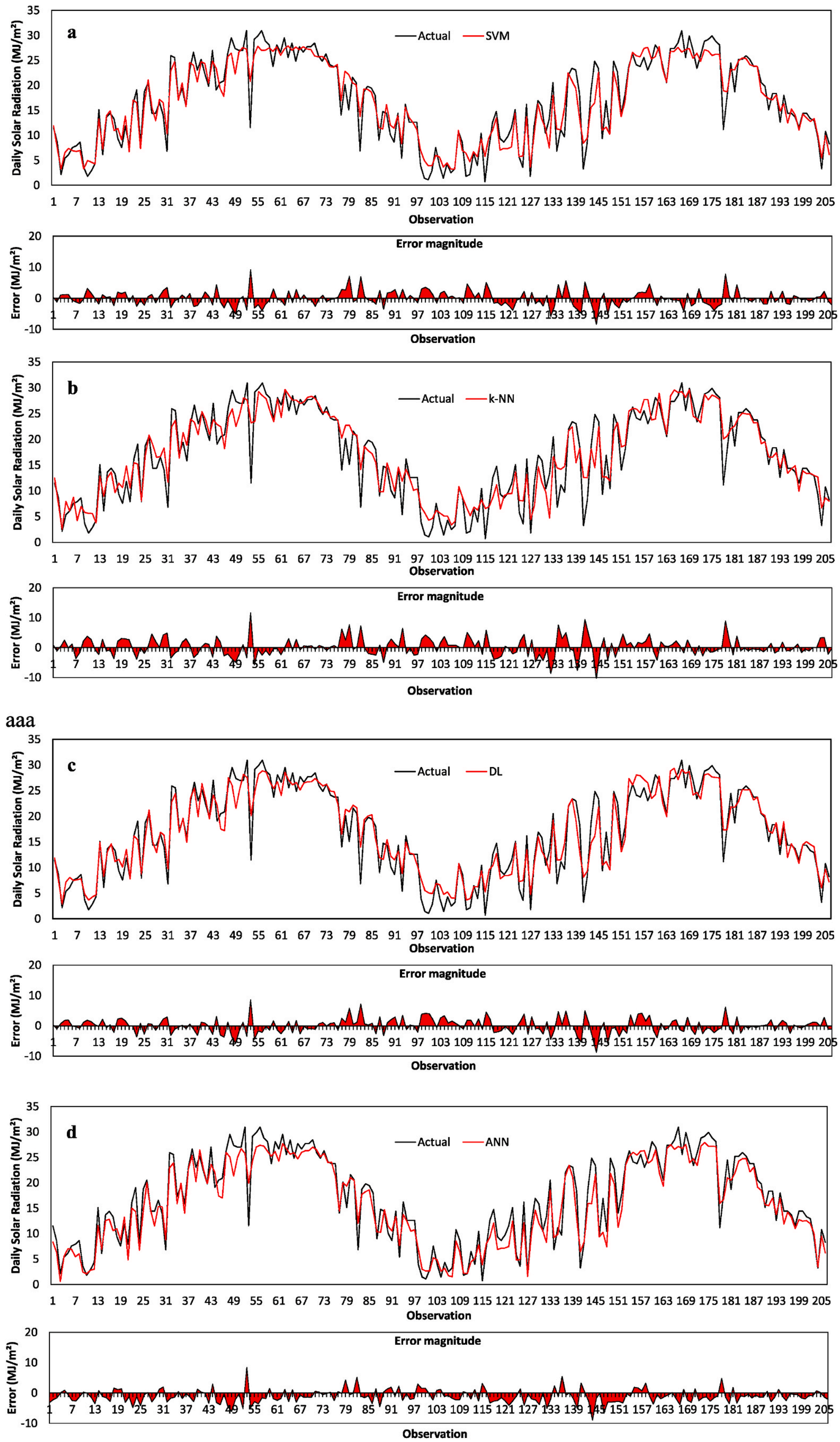


**Fig. 6.** Actual SR data, prediction result and error magnitude for Karaman province a) SVM b) k-NN c) DL, and d) ANN (testing data of 30% for each year).

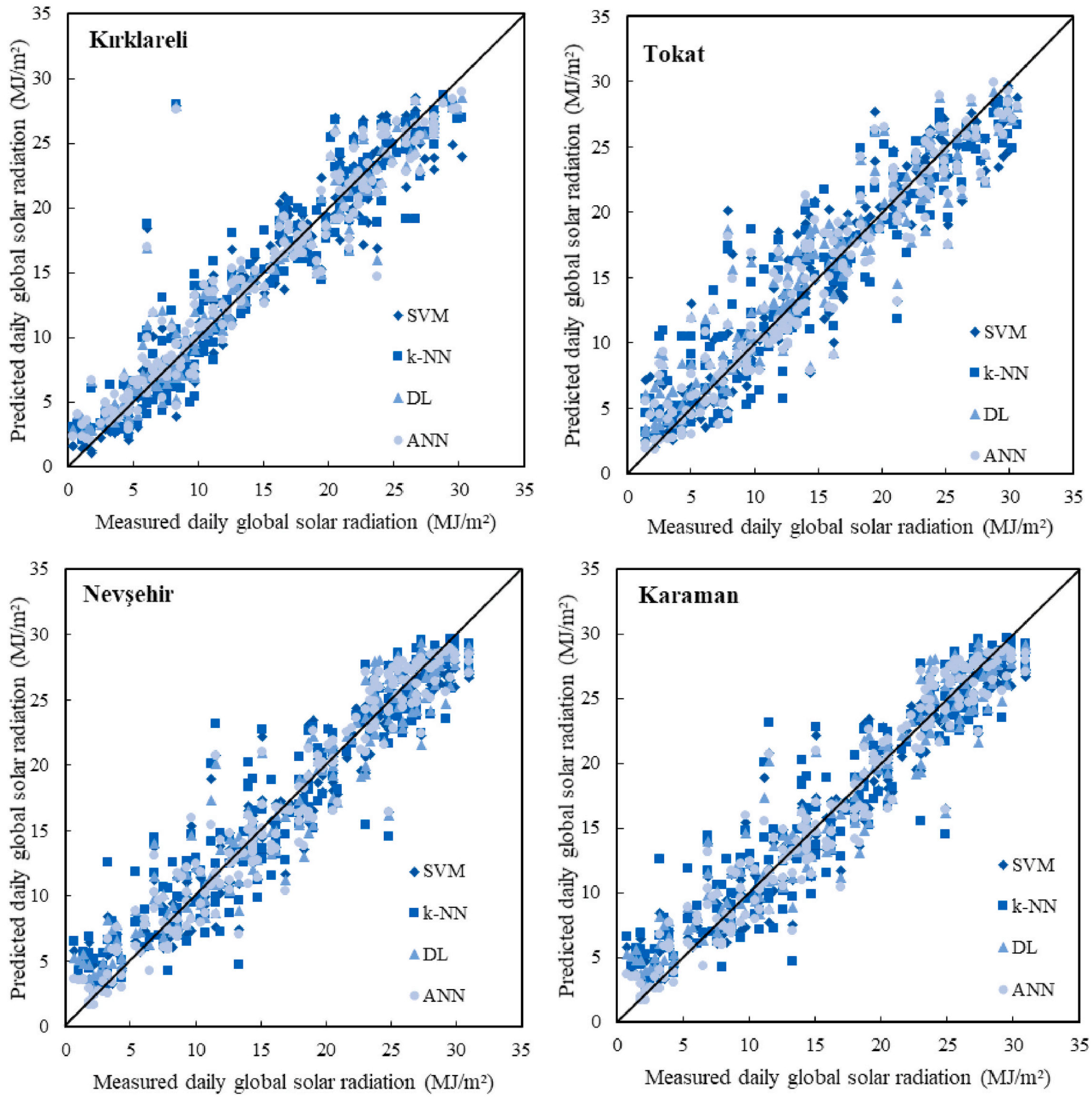


**Fig. 7.** Scatter graph of each city according to machine learning algorithms.

magnitude both measured for the Nevşehir region and predicted by the DL and ANN algorithm, respectively. In terms of $R^2$, both algorithms performed the same result (0.932). However, both algorithms have similar RMSE values. With this viewpoint, even if these two algorithms have similar success, the ANN algorithm comes to the fore when all statistical parameters are considered together. This is because DL algorithm has the highest MBE value (0.676 MJ/m$^2$) among all the algorithms for Nevşehir province (see Table 3). This case makes the DL algorithm an algorithm with worse performance in terms of the t-stat metric. ANN has exhibited better performance in terms of RMSE, MBE, MABE and MAPE metrics. In addition, the t-stat of ANN is meaningful while that of DL is not meaningful. Additionally, the smallest rRMSE value for ANN among all algorithms is seen with 14.10%. Considering the provinces, the smallest rRMSE value is also calculated for Nevşehir.

In Fig. 6a, the original daily global solar radiation of Karaman, in which has the highest global solar radiation potential among all provinces investigated within the present study, and its prediction graphs for SVM are shown. With a general evaluation, it is possible to say that the SVM algorithm is the third most successful algorithm for this province after ANN and DL. Although this algorithm has a good performance (20.32%) in terms of MAPE, it has only been able to pass the k-NN algorithm in terms of the $R^2$, RMSE, rRMSE, MBE and t-stat metrics.

In Fig. 6b, the daily global solar radiation values measured and predicted using the k-NN algorithm are given. As Fig. 6a is analyzed together with Table 3, it is seen that the worst-performing algorithm for Karaman is k-NN algorithm. In terms of MAPE and rRMSE metrics, the highest MAPE (30.24%) and rRMSE (25.19%) values were obtained with k-NN in this province.

Measured and predicted daily global solar radiation values are given in Fig. 6c and d for DL algorithm and ANN algorithm, respectively. When these two algorithms are evaluated together, it can be said that they are the best predictive algorithms for Karaman. However, ANN comes to fore for this province when all statistical metrics were analyzed together.

The scatter plot of the daily average global solar radiation data predicted by using different machine learning algorithms for all four provinces is given in Fig. 7. As seen in the scatter plots, particularly for Tokat province, the error data have wide dispersion. In Nevşehir province, this dispersion was smaller than the other two provinces. The reason why the error dispersion of Tokat is wide may be due to highly changeable of solar radiation data in Tokat and it is in the transition region (See Fig. 1).

Table 4 gives a statistical comparison between the metric results of

**Table 4**
Metric comparison of present paper with the literature studies in the prediction of daily global solar radiation.

| Ref. | Prediction models | Best model | Evaluation metrics | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | MBE (MJ/m$^2$) | RMSE (MJ/m$^2$) | rRMSE (%) | t-stat | MABE (MJ/m$^2$) | $R^2$ | MAPE (%) |
| [86] | SVM, Empirical | SVM | NA | 1.789 | 13.37 | NA | NA | NA | NA |
| [91] | BC, ANN, k-NN | ANN | NA | 3.160 | NA | NA | NA | 0.8600 | NA |
| [90] | SVM-WT, ANN, GP, ARMA | SVM | NA | 1.424 | 7.95 | NA | 0.8405 | 0.9086 | 6.99 |
| [92] | MLP, RBF, GRNN | MLP | NA | 1.940 | NA | NA | NA | 0.8600 | NA |
| [78] | SVR | SVR | NA | 2.004 | 9.03 | NA | 1.252 | 0.9133 | 10.45 |
| [79] | ANFIS, M5Tree, Emprical | ANFIS | NA | 2.070 | NA | NA | NA | 0.9100 | NA |
| [85] | SVR | SVR | NA | NA | NA | NA | NA | 0.9800 | 21.90 |
| [24] | ANFIS, SVM, ANN | SVM | NA | 2.578 | NA | NA | NA | 0.6890 | NA |
| [26] | GEP, ANN, ANFIS, Empirical | ANN | NA | 1.850 | NA | NA | NA | 0.9350 | NA |
| [16] | MEA-ANN, ANN, RF, WNN, Empirical | MEA-ANN | NA | 2.814 | 19.60 | NA | NA | 0.8850 | NA |
| [15] | ANN, Empirical | ANN | 0.384 | 1.855 | NA | NA | NA | 0.92 | NA |
| [80] | ANN, MLR, Empirical | ANN | NA | 3.166 | NA | NA | NA | 0.884 | NA |
| [81] | ANN-ARX | ANN-ARX | NA | 1.730 | 9.90 | NA | 1.040 | 0.870 | 9.45 |
| [82] | KELM | KELM | NA | 2.016 | 11.25 | NA | 1.344 | 0.8203 | NA |
| [83] | FRF-SVM | FRF-SVM | NA | 1.571 | NA | NA | NA | NA | NA |
| [84] | WT-SVM | WT-SVM | NA | 2.317 | 12.57 | NA | NA | NA | 12.67 |
| **Present paper** | **ANN, SVM, DL, k-NN** | **ANN** | **0.195** | **2.157** | **14.10** | **1.280** | **1.597** | **0.9320** | **15.92** |

the present study and previously conducted literature studies regarding the prediction of daily global solar radiation data using various algorithms for various regions. As can be seen in Table 4, the number of preferred metrics in determining the prediction success of algorithms is generally limited. Furthermore, most of the previous studies conducted for the prediction of global solar radiation used the algorithms coming from the same categories. Therefore, statistical metrics have generally presented close results to each other for different algorithms. In this case, the limited number of metrics makes it difficult to determine the prediction success of the algorithms. For example, Aji et al. studied the daily global solar radiation data. The authors found the highest $R^2$ value with 0.9800 in their study. This value is the best one in Table 4. On the other hand, the MAPE value was calculated as 21.90% in the relevant study [85]. This value is also the worst one in Table 4. This case results in a question in the minds. With this viewpoint, seven metrics to comprehensively discuss and determine the algorithms' success are used in the present study.

Another important point to be concluded from Table 4 is that there is not any algorithm always giving the best results for all regions. In other words, even though the same data type is predicted, different algorithms may give the best results region to region. Moreover, there may be a big difference even between the metrics of the same algorithms giving the best results for different regions. For example, Mohammadi et al. achieved the best results in the prediction of daily global solar radiation using the SVM algorithm. In the relevant study, the highest $R^2$ value is calculated as 0.9086 [78]. In another study in which achieving the best prediction results with SVM, Quej et al. found the highest $R^2$ value as 0.6890 [24]. The reason behind these differences may be due to the input variables, dataset size, missing data, feature-selection/extraction, geographical differences, tuning the values of the parameters, etc.

As stated previously sections, the best prediction results are achieved with the ANN algorithm in this paper. In Table 4, as all studies where the best results are achieved with the ANN algorithm are considered, the present paper has mostly presented better metric results in comparison with those of other studies. The reason behind it may be attributed to geographical advantages, dataset diversity, no missing data in the dataset, etc.

## 4. Conclusion

This paper discusses the performance of four different machine learning algorithms (SVM, k-NN, DL and ANN) in the prediction of daily global solar radiation. The study considers various input data (extraterrestrial solar radiation, day length, cloud cover, minimum and maximum temperature are used as attributes) from the four different stations (Kırklareli, Tokat, Nevşehir and Karaman) in Turkey. To evaluate the performance of the machine learning algorithms, seven metrics ($R^2$, RMSE, rRMSE, MBE, MABE, t-stat, and MAPE) are discussed in this study. Then the following conclusions can be drawn based on the present investigation.

1-) Considering the prediction results in terms of $R^2$, it can be said that algorithms used in all provinces have exhibited successful results. $R^2$ values of the algorithms in the study varied between 85.5% and 93.6% depending on the provinces.
2-) The MBE values of the algorithms used in the study were calculated to be mostly positive. MBE has been calculated negatively in Kırklareli and Nevşehir provinces only when using the SVM algorithm. MBE was positive in all other provinces and algorithms.
3-) When all provinces and algorithms are analyzed in terms of MAPE, it is seen that the lowest MAPE value varies between 15.92% and 30.24%. That is, it can be seen that all the prediction results can be categorized as "good prediction" or "reasonable prediction".
4-) As a general evaluation is made in terms of rRMSE, it is seen that rRMSE varies between 14.10% and 25.19% for all provinces and algorithms. Therefore, the prediction success of all algorithms can be evaluated as "good" and "fair" in terms of this metric.
5-) Considering all statistical metrics together for four provinces, the best results are achieved within Nevşehir province.
6-) When a general evaluation is made, it has been seen that DL and ANN algorithms give very close results, particularly considering the $R^2$ and RMSE metrics. Therefore, different parameters should be taken into consideration to discuss and decide on the success of these algorithms.
7-) The most important parameter that distinguishes DL and ANN algorithms is t-stat. Even if these two algorithms are very close to each other particularly in Tokat and Nevşehir provinces, the DL algorithm has exceeded the t-critical value both for these two provinces. This is due to the high MBE values that the DL algorithm has. Therefore, ANN has been proposed as a more successful algorithm for these provinces.
8-) Considering the error magnitudes of the observations randomly determined in the study, it is seen that the error magnitudes in the use of ANN algorithm are very low in comparison with those of other algorithms, particularly in the observations where solar radiation is lower.

## Credit author

Ümit Ağbulut: Conceptualization, Investigation, Formal analysis, Resources, Writing - original draft, Writing - review & editing. Ali Etem Gürel: Conceptualization, Investigation, Formal analysis, Resources, Writing - original draft, Writing - review & editing. Yunus Biçen: Conceptualization, Investigation, Formal analysis, Resources, Writing - original draft, Writing - review & editing.

## Declaration of competing interest

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

## Acknowledgements

This work is supported by Düzce University Scientific Research Projects Coordination Unit. Project Number: 2019.07.04.1049.2019.07.04.1049.

## References

[1] Khanlari A, Sözen A, Şirin C, Tuncer AD, Gungor A. Performance enhancement of a greenhouse dryer: analysis of a cost-effective alternative solar air heater. J Clean Prod 2020;251:119672.
[2] Kalogirou SA. Designing and modeling solar energy systems. second ed. Solar Energy Engineering; 2014. p. 583–699 [Chapter 11].
[3] Khanlari A, Sözen A, Afshari F, Şirin C, Tuncer AD, Gungor A. Drying municipal sewage sludge with v-groove triple-pass and quadruple-pass solar air heaters along with testing of a solar absorber drying chamber. Sci Total Environ 2020;709:136198.
[4] Yıldırım HB, Çelik Ö, Teke A, Barutçu B. Estimating daily Global solar radiation with graphical user interface in Eastern Mediterranean region of Turkey. Renew Sustain Energy Rev 2018;82:1528–37.
[5] Yadav AK, Malik H, Chandel SS. Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models. Renew Sustain Energy Rev 2014;31:509–19.
[6] Balog I, Podrascanin Z, Spinelli F, Caputo G, Siviero R, Benedetti A. July). Hourly forecast of solar radiation up to 48h with two runs of weather research forecast model over Italy. AIP conference proceedings, vol. 2126. AIP Publishing LLC; 2019. p. 190004. No. 1.
[7] 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.
[8] Dong N, Chang JF, Wu AG, Gao ZK. A novel convolutional neural network framework based solar irradiance prediction method. Int J Electr Power Energy Syst 2020;114:105411.
[9] Jiang Y, Long H, Zhang Z, Song Z. Day-ahead prediction of bihourly solar radiance with a Markov switch approach. IEEE Transactions on Sustainable Energy 2017;8(4):1536–47.
[10] Gürel AE, Ağbulut Ü, Biçen Y. Assessment of machine learning, time series, response surface methodology and empirical models in prediction of global solar radiation. J Clean Prod 2020:122353.
[11] TSMS. Turkish State Meteorological Service. January 7, 2020. https://mgm.gov.tr/eng/.
[12] Zang H, Xu Q, Bian H. Generation of typical solar radiation data for different climates of China. Energy 2012;38(1):236–48.
[13] Fan J, Wang X, Wu L, Zhang F, Bai H, Lu X, Xiang Y. 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 2018;156:618–25.
[14] Kisi O, Parmar KS. Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. J Hydrol 2016;534:104–12.
[15] Jahani B, Mohammadi B. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. Theor Appl Climatol 2019;137(1–2):1257–69.
[16] Feng Y, Gong D, Zhang Q, Jiang S, Zhao L, Cui N. Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation. Energy Convers Manag 2019;198:111780.
[17] Khosravi A, Nunes RO, Assad MEH, Machado L. Comparison of artificial intelligence methods in estimation of daily global solar radiation. J Clean Prod 2018;194:342–58.
[18] Sharifi SS, Rezaverdinejad V, Nourani V. Estimation of daily global solar radiation using wavelet regression, ANN, GEP and empirical models: a comparative study of selected temperature-based approaches. J Atmos Sol Terr Phys 2016;149:131–45.
[19] Jiang Y. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. Energy Pol 2008;36(10):3833–7.
[20] Huang C, Zhao Z, Wang L, Zhang Z, Luo X. Point and interval forecasting of solar irradiance with an active Gaussian process. IET Renew Power Gener 2020;14(6):1020–30.
[21] Long H, Zhang Z, Su Y. Analysis of daily solar power prediction with data-driven approaches. Appl Energy 2014;126:29–37.
[22] Liu Y, Zhou Y, Chen Y, Wang D, Wang Y, Zhu Y. 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 2020;146:1101–12.
[23] Bayrakçı HC, Demircan C, Keçebaş A. The development of empirical models for estimating global solar radiation on horizontal surface: a case study. Renew Sustain Energy Rev 2018;81:2771–82.
[24] Quej VH, Almorox J, Arnaldo JA, Saito L. ANFIS, SVM and ANN soft-computing techniques to estimate daily global solar radiation in a warm sub-humid environment. J Atmos Sol Terr Phys 2017;155:62–70.
[25] Marzo A, Trigo-Gonzalez M, Alonso-Montesinos J, Martínez-Durbán M, López G, Ferrada P, Batlles FJ. Daily global solar radiation estimation in desert areas using daily extreme temperatures and extraterrestrial radiation. Renew Energy 2017;113:303–11.
[26] Mehdizadeh S, Behmanesh J, Khalili K. Comparison of artificial intelligence methods and empirical equations to estimate daily solar radiation. J Atmos Sol Terr Phys 2016;146:215–27.
[27] Tymvios FS, Jacovides CP, Michaelides SC, Scouteli C. Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation. Sol Energy 2005;78(6):752–62.
[28] Kaba K, Sarıgül M, Avcı M, Kandırmaz HM. Estimation of daily global solar radiation using deep learning model. Energy 2018;162:126–35.
[29] Marzouq M, Bounoua Z, El Fadili H, Mechaqrane A, Zenkouar K, Lakhliai Z. New daily global solar irradiation estimation model based on automatic selection of input parameters using evolutionary artificial neural networks. J Clean Prod 2019;209:1105–18.
[30] MENR. Republic of Turkey ministry of energy and natural resources january, vol. 7; 2020 [Online], http://www.yegm.gov.tr/MyCalculator/.
[31] Bakirci K. Correlations for estimation of daily global solar radiation with hours of bright sunshine in Turkey. Energy 2009;34(4):485–501.
[32] Methaprayoon Kittipong, et al. An integration of ANN wind power estimation into unit commitment considering the forecasting uncertainty. IEEE Trans Ind Appl 2007;43(6):1441–8.
[33] Mabel M Carolin, Fernandez Eugene. Analysis of wind power generation and prediction using ANN: a case study. Renew Energy 2008;33(5):986–92.

[34] Palani Sundarambal, , Shie-Yui Liong, Tkalich Pavel. An ANN application for water quality forecasting. Mar Pollut Bull 2008;56(9):1586–97.
[35] Behrang MA, et al. The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data. Sol Energy 2010;84(8):1468–80.
[36] Azlah, Firdaus Muhammad Azfar, et al. Review on techniques for plant leaf classification and recognition. Computers 2019;8(4):77.
[37] Liu Hongyu, Lang Bo. Machine learning and deep learning methods for intrusion detection systems: a survey. Appl Sci 2019;9(20):4396.
[38] Samur, Iltuzer Zeynep, Gül Tekin Temur. The use of artificial neural network in option pricing: the case of S&P 100 index options. World Academy of Science, Engineering and Technology 2009;54:326–31.
[39] Ramil A, et al. A computer vision system for identification of granite-forming minerals based on RGB data and artificial neural networks. Measurement 2018; 117:90–5.
[40] Chen Hui-Ling, et al. An efficient diagnosis system for detection of Parkinson's disease using fuzzy k-nearest neighbor approach. Expert Syst Appl 2013;40(1): 263–71.
[41] Hu Li-Yu, et al. The distance function effect on k-nearest neighbor classification for medical datasets. SpringerPlus 2016;5(1):1–9.
[42] Tan Songbo. Neighbor-weighted k-nearest neighbor for unbalanced text corpus. Expert Syst Appl 2005;28(4):667–71.
[43] Zhang Shichao, et al. Efficient knn classification with different numbers of nearest neighbors. IEEE transactions on neural networks and learning systems 2017;29(5): 1774–85.
[44] Rodrigues Érick O. Combining Minkowski and Cheyshev: new distance proposal and survey of distance metrics using k-nearest neighbours classifier. Pattern Recogn Lett 2018;110:66–71.
[45] Myhre Jonas Nordhaug, et al. Robust clustering using a kNN mode seeking ensemble. Pattern Recogn 2018;76:491–505.
[46] Taşcı Erdal, Onan Aytuğ. Investigation of the effect of K-nearest neighbor algorithm parameters on classification performance. In: Turkish) akademik bilişim; 2016.
[47] Maillo Jesus, et al. Exact fuzzy k-nearest neighbor classification for big datasets. In: IEEE international conference on fuzzy systems. FUZZ-IEEE). IEEE; 2017. 2017.
[48] Saikia Jyotishman, et al. K-nearest neighbor hardware accelerator using in-memory computing SRAM. In: IEEE/ACM international symposium on low power electronics and design (ISLPED). IEEE; 2019. 2019.
[49] Zendehboudi Alireza, Baseer MA, Saidur R. Application of support vector machine models for forecasting solar and wind energy resources: a review. J Clean Prod 2018;199:272–85.
[50] Chapelle Olivier, et al. Choosing multiple parameters for support vector machines. Mach Learn 2002;46:1–3. 131-159.
[51] Onel Melis, et al. Big data approach to batch process monitoring: simultaneous fault detection and diagnosis using nonlinear support vector machine-based feature selection. Comput Chem Eng 2018;115:46–63.
[52] Park Hyejin, Kim Namhyoung, Lee Jaewook. Parametric models and non-parametric machine learning models for predicting option prices: empirical comparison study over KOSPI 200 Index options. Expert Syst Appl 2014;41(11): 5227–37.
[53] Zhang ChunHua, Tian YingJie, Deng NaiYang. The new interpretation of support vector machines on statistical learning theory. Sci China, Ser A: Mathematics 2010; 53(1):151–64.
[54] Kim Sungdo, Mun Byeong Min, Bae Suk Joo. Data depth based support vector machines for predicting corporate bankruptcy. Appl Intell 2018;48(3):791–804.
[55] Yoon Min, Yun Yeboon, Nakayama Hirotaka. A role of total margin in support vector machines. In: Proceedings of the international joint conference on neural networks, vol. 3. IEEE; 2003. 2003.
[56] Birzhandi Pardis, et al. Reduction of training data using parallel hyperplane for support vector machine. Appl Artif Intell 2019;33(6):497–516.
[57] Najafabadi Maryam M, et al. Deep learning applications and challenges in big data analytics. Journal of Big Data 2015;2(1):1.
[58] Wang Chao, et al. DLAU: a scalable deep learning accelerator unit on FPGA. IEEE Trans Comput Aided Des Integrated Circ Syst 2016;36(3):513–7.
[59] Bharati Aparna, et al. Detecting facial retouching using supervised deep learning. IEEE Trans Inf Forensics Secur 2016;11(9):1903–13.
[60] Weston Jason, et al. "Deep learning via semi-supervised embedding." Neural networks: tricks of the trade. Berlin, Heidelberg: Springer; 2012. p. 639–55.
[61] Miyato Takeru, et al. Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Trans Pattern Anal Mach Intell 2018;41(8):1979–93.
[62] Chen Yize, Wang Xiyu, Zhang Baosen. An unsupervised deep learning approach for scenario forecasts. In: Power systems computation conference (PSCC). IEEE; 2018. 2018.
[63] Fan Cheng, et al. Deep learning-based feature engineering methods for improved building energy prediction. Appl Energy 2019;240:35–45.
[64] Mei Shaohui, et al. Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks. IEEE Trans Geosci Rem Sens 2017;55(8):4520–33.
[65] Lin Yi-zhou, Nie Zhen-hua, Ma Hong-wei. Structural damage detection with automatic feature-extraction through deep learning. Comput Aided Civ Infrastruct Eng 2017;32(12):1025–46.
[66] Hua Yuming, Guo Junhai, Zhao Hua. Deep belief networks and deep learning. In: Proceedings of 2015 international conference on intelligent computing and internet of things. IEEE; 2015.
[67] Morgan. Nelson. "Deep and wide: multiple layers in automatic speech recognition. IEEE Trans Audio Speech Lang Process 2011;20(1):7–13.
[68] Chen Yushi, et al. Deep learning-based classification of hyperspectral data. IEEE Journal of Selected topics in applied earth observations and remote sensing 2014;7 (6):2094–107.
[69] Zhou Lina, et al. Machine learning on big data: opportunities and challenges. Neurocomputing 2017;237:350–61.
[70] Zang H, Cheng L, Ding T, Cheung KW, Wang M, Wei Z, Sun G. Application of functional deep belief network for estimating daily global solar radiation: a case study in China. Energy 2020;191:116502.
[71] Heinemann AB, Van Oort PA, Fernandes DS, Maia ADHN. Sensitivity of APSIM/ ORYZA model due to estimation errors in solar radiation. Bragantia 2012;71(4): 572–82.
[72] Tuncer AD, Sözen A, Afshari F, Khanlari A, Şirin C, Gungor A. Testing of a novel convex-type solar absorber drying chamber in dehumidification process of municipal sewage sludge. J Clean Prod 2020;272:122862.
[73] Rehman S. Solar radiation over Saudi Arabia and comparisons with empirical models. Energy 1998;23(12):1077–82.
[74] Yang L, Cao Q, Yu Y, Liu Y. Comparison of daily diffuse radiation models in regions of China without solar radiation measurement. Energy 2020;191:116571.
[75] Gouda SG, Hussein Z, Luo S, Yuan Q. Model selection for accurate daily global solar radiation prediction in China. J Clean Prod 2019;221:132–44.
[76] Ceylan I, Gürel AE, Ergün A. The mathematical modeling of concentrated photovoltaic module temperature. Int J Hydrogen Energy 2017;42(31):19641–53.
[77] Emang D, Shitan M, Abd Ghani AN, Noor KM. Forecasting with univariate time series models: A case of export demand for peninsular Malaysia's moulding and chipboard. J. Sustain. Develop. 2010;3(3):157–61.
[78] Mohammadi K, Shamshirband S, Anisi MH, Alam KA, Petković D. Support vector regression based prediction of global solar radiation on a horizontal surface. Energy Convers Manag 2015;91:433–41.
[79] Wang L, Kisi O, Zounemat-Kermani M, Zhu Z, Gong W, Niu Z, Liu Z. Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree. Int J Climatol 2017;37(3):1141–55.
[80] Antonopoulos VZ, Papamichail DM, Aschonitis VG, Antonopoulos AV. Solar radiation estimation methods using ANN and empirical models. Comput Electron Agric 2019;160:160–7.
[81] Shamshirband S, Mohammadi K, Piri J, Petković D, Karim A. Hybrid auto-regressive neural network model for estimating global solar radiation in Bandar Abbas, Iran. Environmental Earth Sciences 2016;75(2):172.
[82] Shamshirband S, Mohammadi K, Chen HL, Samy GN, Petković D, Ma C. Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: a case study for Iran. J Atmos Sol Terr Phys 2015;134:109–17.
[83] Baser F, Demirhan H. A fuzzy regression with support vector machine approach to the estimation of horizontal global solar radiation. Energy 2017;123:229–40.
[84] Deo RC, Wen X, Qi F. A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. Appl Energy 2016;168:568–93.
[85] Aji P, Wakamori K, Mineno H. 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; 2018. p. 1–5.
[86] Chen JL, Li GS, Wu SJ. Assessing the potential of support vector machine for estimating daily solar radiation using sunshine duration. Energy Convers Manag 2013;75:311–8.
[87] Fan J, Wu L, Zhang F, Cai H, Ma X, Bai H. Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China. Renew Sustain Energy Rev 2019; 105:168–86.
[89] Meenal R, Selvakumar AI. Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters. Renew Energy 2018;121:324–43.
[90] Mohammadi K, Shamshirband S, Tong CW, Arif M, Petković D, Ch S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manag 2015;92:162–71.
[91] Moreno A, Gilabert MA, Martínez B. Mapping daily global solar irradiation over Spain: a comparative study of selected approaches. Sol Energy 2011;85(9): 2072–84.
[92] Wang L, Kisi O, Zounemat-Kermani M, Salazar GA, Zhu Z, Gong W. Solar radiation prediction using different techniques: model evaluation and comparison. Renew Sustain Energy Rev 2016;61:384–97.