Journal of Renewable
Energy and Environment

Journal Homepage: www.jree.ir

Materials and Energy Research Center
MERC

Research Note

Predicting Solar Power Generation Based on the Combination of Meteorological Parameters in Iran: Neural Networks Approach

Abbas Ahmadi [a], Mahsa Zaman [b], Siab Mamipour [b*]

[a] *Department of Industrial Engineering & Management Systems, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran.*
[b] *Department of Economics, Kharazmi University, Tehran, Iran.*

PAPER INFO

Paper History:
Received: 07 October 2022
Revised: 03 January 2023
Accepted: 17 January 2023

Keywords:
Solar Power Plants,
Metrological Parameters,
Artificial Neural Networks,
Iran

ABSTRACT

Clean solar energy is one of the best sources of energy. Solar power plants can generate electricity in Iran due to their large number of sunny days. This paper presents a short-term forecasting approach based on artificial neural networks (ANNs) for selected solar power plants in Iran and ranks the input variables of the neural network according to their importance. Two solar power plants in Hamadan province (Amirkabir and Khalij-Fars) were selected for the project. The output of solar power plants is dependent on weather conditions. Solar radiation on the horizontal plane, air temperature, air pressure, day length, number of sunny hours, cloudiness, and airborne dust particles are considered input variables in this study to predict solar power plant output. Forecasting model selection is based on considering zero and nonzero quantities of target variables. The results show that solar production forecasting based on meteorological parameters in the Khalij-Fars is more accurate than Amirkabir. The global solar radiation, air temperature, number of sunny hours, day length, airborne dust particles, cloudiness, air pressure, and dummy variables[1] are the order of the most important inputs to solar power generation. Results show simultaneous influences of radiation and temperature on solar power plant production.

https://doi.org/10.30501/jree.2023.363386.1461

[1]. The first half of the year is counted as one, and the second half is counted as zero.

1. INTRODUCTION

Energy production has long been accompanied by carbon emissions. Carbon emission is one of the most important environmental issues leading communities to use renewable energy sources. According to the EIA report of Country Analysis Executive summary, Iran was the fifth-largest oil producer in 2020 and the third-largest gas producer in 2019.

Iran is ranked the third as the world's largest proved reserve holder of oil and second-largest proved reserve holder of natural gas. Easy access to fossil fuels means that 73 % of Iran's net electricity generation is from natural gas, 15 % from oil, 10 % from hydropower, 2 % from nuclear power plants, and just less than 1 % from coal and non-hydro renewable energies (EIA, 2021). There are several major oil and gas refineries in Iran, which have led to the country ranking seventh in carbon dioxide emissions worldwide indicating the importance of reducing them (Mamipour et al., 2019).

In spite of the cheaper cost of production of electricity from the combustion of fossil fuels and the lower amount of electricity produced by renewable power plants and fossil fuel power plants, the development of renewable energy is more pragmatic for environmental reasons. Since the establishment of new solar power plants in Iran is assigned to the private sector, the return of capital to investors is an important aspect and it is indicative that accurate predictions are critical to the establishment of new solar power plants. Some deficiencies might dampen investment enthusiasm for building new solar power plants. To maximize the amount of power generated in a given region, the meteorological variables affecting solar power output must be precisely calculated. A weather forecast can help investors select the optimal equipment and panels for a particular region. Due to Iran's vast size and wide range of climates, this factor is very important in choosing the best sites for establishing new power plants. An accurate assessment of the impact of weather variables on the output of solar power plants can greatly affect the optimum choice for their siting.

The general idea is that the high radiation points are the best regions for establishing solar power plants, but the temperature will rise as the sun rises, which will decrease the efficiency of the power plant (Bhavani et al., 2021). The main factors that affect the performance of power plants should be ranked empirically based on climate and region. In determining the output of solar power plants, only the factors

*Corresponding Author's Email: s.mamipoor@khu.ac.ir (S. Mamipour)
URL: https://www.jree.ir/article_166555.html

Please cite this article as: Ahmadi, A., Zaman, M. & Mamipour, S. (2023). Predicting solar power generation based on the combination of meteorological parameters in Iran: Neural networks approach. *Journal of Renewable Energy and Environment (JREE)*, *10*(4), 131-145. https://doi.org/10.30501/jree.2023.363386.1461.

that are most likely to affect solar power production should be taken into account.

As shown in Figure 1, solar power plants should be established in flat, wide areas that are near the main power grids. To ensure that the selected place is a suitable location concerning points in its immediate vicinity, one needs to examine and analyze the most important meteorological parameters of the selected region.

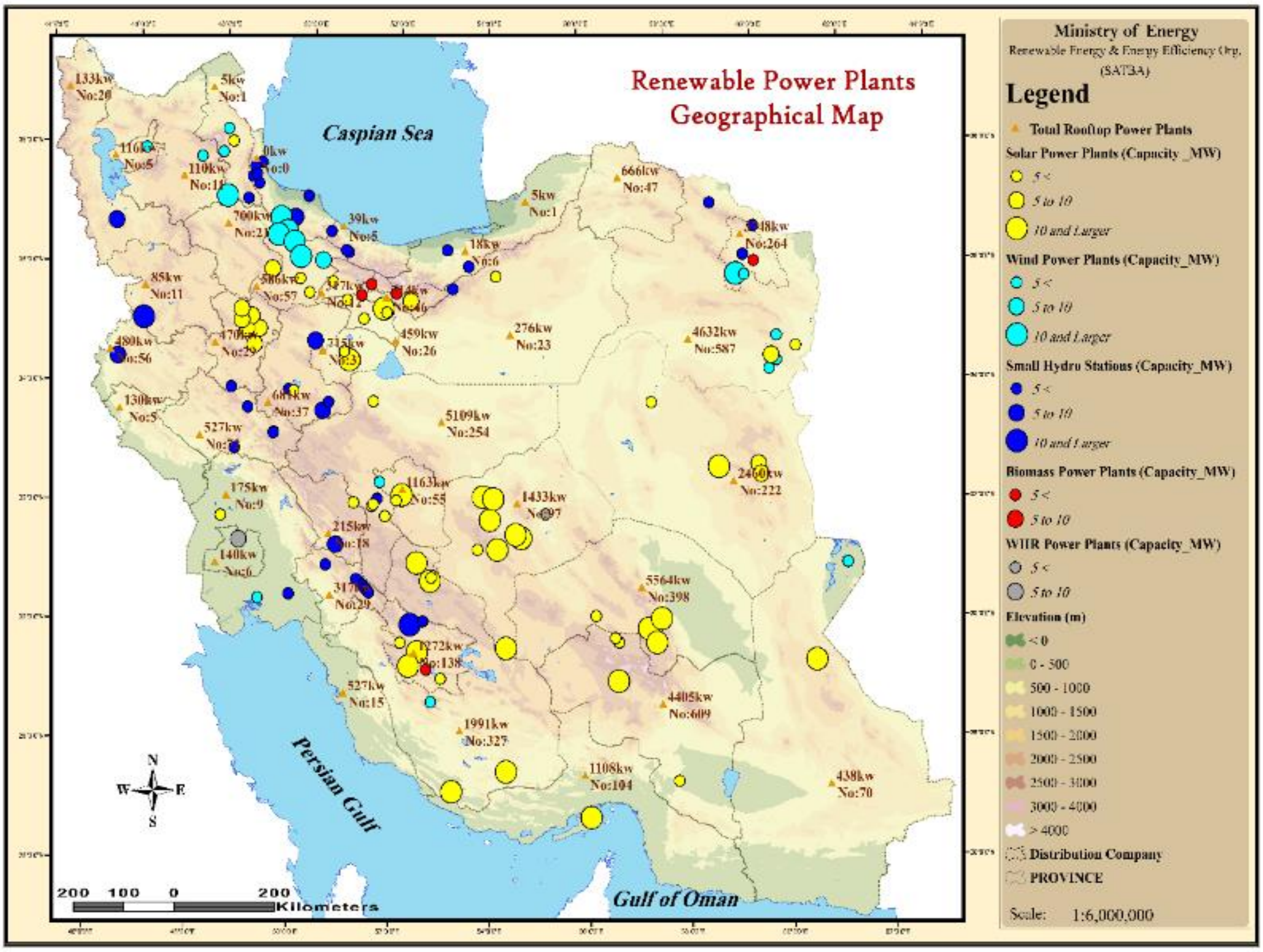


Figure 1. Overview of Iran's renewable power plants –end of year summary 2020

Figure 1 outlines the location and capacity of installed renewable power plants in Iran. The north of Iran has no solar power plants in operation because the weather condition is mild and there are more rainy days than sunny days.

As the sun shines for long periods in most regions of Iran, this clean energy may be used to generate electricity. As a result of intermittent fluctuations in PV[3] system output, production is also volatile during the day. Some particles in the atmosphere, such as water vapor and gases in the atmosphere, absorb some of the sun's rays, while others, such as dust particles in the atmosphere, disperse this energy into space (Carra et al., 2018). In general, about 340 watts per square meter of solar energy falls to the earth, but only 48 percent of it reaches the ground and can be used for solar energy production. 29 % of solar radiation is reflected in space by clouds, bright surfaces, and atmospheres. Gases in the atmosphere, dust, and other particles absorb 23 percent of this energy (NASA, 2009). Approximately, 48 % of this solar radiation can be used by photovoltaic panels to produce solar energy. Accordingly, the amount of sunlight reaching the earth's surface fluctuates with climate conditions, Thus, photovoltaic cells have variable output. Atmospheric fluctuations affect solar electricity production. There have been many studies attempting to model the output of solar energy production based on this fluctuation, which has led to numerous studies estimating the output of solar plants (Jung et al., 2020; Vaka & Talukdar, 2020; Zaaoumi et al., 2021; Zhao et al., 2021). The major categories for predicting solar energy are theoretical sunshine-based models, empirical meteorological parameters, and combinations of both meteorological models and sunshine-based models (Jahani et al., 2017). It has been reported that the amount of global solar radiation on the horizontal plane, air temperature, air pressure, number of sunny hours, cloudiness, dust particles in the air, and relative humidity can have an impact on the performance of solar power plants (Bugała et al., 2018; Khosravi et al., 2018; Loghmari et al., 2018; Rao et al., 2018).

In some cases, solar power plants are not located near meteorological stations. However, they need to measure meteorological parameters to accurately predict power plant output. Since establishing new meteorological stations is very expensive, this study investigates what can be done to make accurate predictions more reasonable.

2. LITERATURE REVIEW

[3] A photovoltaic (PV) system is made up of one or more solar panels with an inverter and other electrical and mechanical equipment that converts sunlight into electricity.

The following papers (Bugała et al., 2018; Ghritlahre & Prasad, 2018; Loghmari et al., 2018; Olden et al., 2004; Rao et al., 2018; Shireen et al., 2018) discuss the application of neural networks to solar system prediction. In previous research, neural networks have been used to estimate the following topics:

i. Solar energy prediction
ii. Solar radiation prediction
iii. Predicting the output of solar systems
iv. Meteorological ANN models for Iran's weather condition

Detailed in the topics below, it explains that the connection weight method is selected in the current research based on previous studies to analyze the importance of input variables.

i. Solar energy prediction

Shireen et al. (2018) developed a model based on repeated multi-purpose learning. Owners of solar power systems can benefit from modeling PV output time series because this allows them to understand how energy systems behave over time. An effective method of multifunctional learning is proposed for the MTL-GP-TS time series to predict PV output. By combining PV measurements from multiple solar panels with similar traits, measurements are improved. Learning the proposed MTL-GP-TS model iteratively uncovers hidden or missing values in a set of panel-related time series that are potentially useful in predicting PV trends. Furthermore, it improves the traditional multifunctional learning process and generalizes the Gaussian process of learning both global trends and irregular local components. Based on a real-world case study, the proposed approach can improve conventional approaches significantly.

ii. Solar radiation prediction

Halabi et al. (2018) evaluated the performance of hybrid models of adaptive neuro-fuzzy inference systems for predicting monthly solar radiation. The output of solar energy systems is highly dependent on solar radiation. Therefore, accurate forecasting of solar radiation is very important. Hence, a consistent independent fuzzy inference system and a hybrid model were developed to predict monthly solar radiation following various meteorological parameters such as irradiation time S(h) and air temperature. Their proposed hybrid models include particle swarm optimization, genetic algorithms, and differential evaluation. To evaluate the capability and efficiency of the proposed model, several statistical indicators such as mean squared error, correlation coefficient, and mean absolute error are used. Performance evaluation over various statistical indicators exhibited a high correlation for all of the developed modules. Hybrid particle swarm optimization has obtained the best statistical indicators in all of the models. An accurate comparison with other studies has been performed to validate the accuracy of the proposed prediction models and their appropriateness. The results showed that the developed hybrid models had the highest reliability, more accurate estimation, and the most efficient methods for global prediction.

Loghmari et al. (2018) compared the performance of two models of global solar radiation. They developed two global satellite models for solar radiation: an artificial neural network (ANN) and a reverse weighting model (IDW). The goal is to predict global solar radiation at a distance of more than 50 km. The ANN model uses meteorological data in the inventory target area, while the IDW model employs global solar radiation measured in neighboring areas. For the construction and validation of the models, for 5 consecutive years (2008-2012), the values of 5 different meteorological parameters were collected monthly from 10 meteorological stations located in the south and center of Tunisia. The evaluation results of the two models provide comparable results. For the developed ANN model, the average root of the mean square error is 6.4 %, while for the IDW model, it is 5.11 %. The IDW model is simpler and slightly more accurate than the ANN model. This study examined the behavior of two models for different climate conditions through two scenarios. The results show that the number of samples that ANN is trained to predict Global Horizontal Irradiance (GHI) is more important than the climatic conditions from which these samples are retrieved. However, providing input data from sites with similar weather conditions to the predicted area increases the accuracy of the IDW model. In the present study, data from meteorological stations are collected because, in real conditions, all countries cannot have free access to satellite meteorological data.

Rao et al. (2018) analyzed different combinations of meteorological parameters in predicting the amount of total horizontal solar energy radiation with a neural network approach. For the input variables of biennial data for the characteristics of total daily radiation, minimum temperature, maximum temperature, minimum and maximum temperature difference, sunny hours, sunny hours in theory, and extraterrestrial radiation are considered. Different combinations of input variables were considered to predict monthly solar radiation. Out of 32 possible modes, models with a combination of the theoretical sunny hours and extraterrestrial radiation had the best performance. These two parameters are available for any location and do not need to be measured. The best performance belonged to the case with the least number of inputs. In the present study, we try to eliminate unnecessary parameters and unreasonable combinations from the predicted models and with fewer computations try to get more accurate results.

iii. Predicting the output of solar systems

Short-term predictions of power generation in photovoltaic systems were made (Bugała et al., 2018). An in-depth analysis of the input data measured in Poland showed that the effect of some variables such as air pressure and day length was statistically insignificant. The values of skewness, elongation, and the results of experiments applied to investigate the distribution of the dependent variable for daily power generation indicated that the linear regression model should not be the only method in the forecasting process. The developed neural network was based on the RBF model with a quality test of approximately 93 % and an RMS error of 0.02 %. The input variables required for the proposed ANN model included the number of sunshine hours, day length, air pressure, maximum air temperature, amount of daily radiation, and cloudiness.

Ghritlahre & Prasad (2018) applied neural network techniques to predicting the performance of solar collectors. Solar collectors are designed for low- to medium-temperature ranges. Therefore, the optimal design of collector systems helps to increase solar energy efficiency. In this research, a neural network technique is proposed to estimate the thermal performance of the multilateral flow of a porous solar air heating bed. This study further addresses the research gap in research conducted on solar collectors. The present study has confirmed that both temperature and global horizontal solar radiation in real case studies are simultaneously significant to the output of solar power plants.

iv. Meteorological ANN models for Iran's weather condition

Gorjian et al. (2015) modeled solar radiation potential in Iran based on the meteorological and geological data of 31 stations spreading all over the country. They considered solar radiation as the target variable and month of the year, latitude, longitude, altitude, sunshine duration, minimum air temperature, maximum air temperature, maximum daily earth temperature, atmospheric pressure, and precipitation as input variables.

Solar radiation reaching the Earth was modelled using ANFIS, NN-ARX (Piri & Kisi, 2015). It was a case study of two synoptic stations of Zahedan and Bojnurd. The data included sunshine hours, maximum and minimum temperatures, average relative humidity, and solar radiation. A comparison was made between artificial intelligence models and empirical models. It was found that ANFIS performed better than the empirical models in estimating daily solar radiation.

Khosravi et al. (2018) performed hourly predictions of solar radiation on Abu Musa Island using machine learning algorithms. This study proposed machine learning algorithms for predicting hourly solar radiation. Prediction models were developed based on two types of input data. The first model uses local time, temperature, pressure, wind speed, and relative humidity as input variables of the model (N1), and the second model predicted solar radiation (N2). Predictive models use only past solar radiation values to estimate future values. For this purpose, a multilayer feed-forward neural network (MLFFNN), radial basis function neural network (RBFNN), support vector regression (SVR), fuzzy inference system (FIS), and an adaptive fuzzy inference system (ANFIS) were used. The results showed that for N1 models, SVR and MLFFNN had the maximum predicted solar radiation performance with R = 0.9999 and 0.9957, respectively. For N2, SVR, MLFFNN, and ANFIS models reported a correlation coefficient more than 0.95 for the test data set.

Many empirical studies predict solar radiation depending on the weather conditions in Iran, but none of these studies has investigated the impact of meteorological parameters on the production of solar power plants, and their focus is only on prediction solar radiation. Obviously, to predict the production of solar power plants based on meteorological parameters, real and accurate data are required, which is followed in this research. Therefore, the main contribution of this paper is that it tries to predict the output of solar power plants by using meteorological parameters. In this study, eight meteorological parameters that affect the output of solar power plants are considered as input variables. In meteorological and hydrological research, the question always arises as to which one of the input variables of the neural network has a more important role in prediction. Although model sensitivity analysis in most studies is not common, there are methods to determine the importance of input variables. Garson methods, connection weights, partial derivatives, sensitivity analysis, adding a parameter to the model, and removing a parameter from the model are some of the methods for measuring the importance of parameters (Olden et al., 2004). Based on the research of Olden et al. (2004), the best method is the connection weights. In this study, the connection weights method is used to rank the importance of input variables.

3. METHODOLOGY

ANN models are the most common data mining models inspired by human brain functions and are used to model both linear and nonlinear systems. In the present study, data with three-hour and daily frequency for the meteorological parameters are considered from the IRIMO organization for the parameters of cloudiness, temperature, air pressure, and the number of sunny hours, but the purpose of the study includes hourly and daily prediction. Consequently, hourly and daily data sets were retrieved from the SODA website for temperature, air pressure, solar radiation, day length, and global horizontal irradiation. Since the website does not report cloudiness information and the number of sunny hours, data from the IRIMO organization was used. By assuming that the amount of cloudiness and the number of sunny hours during three hours are constant, these two variables are converted into hourly ones. One of the highly correlated variables of meteorological features is cloudiness. This variable is often reported at airport meteorological stations. In the city of Qahavand, cloudiness was not reported for 5 months; therefore, cloudiness data were retrieved from the nearest meteorological station, Hamadan airport station. ANNs were made of three layers. The first one is called the input layer which consists of input variables introduced in Table 1.

Table 1. Input and output parameters

Input Output	Symbols	Unit	Frequency	Source	Explanation
Input	GHI	Wh/m^2	Hourly & daily	SODA (Professionals, 2021)	Global solar radiation on a horizontal plane at the ground level
	Temp	Celsius degree[4]	Hourly & daily	SODA	Temperature at a height of 2 meters above the ground
	Press	hPa 5	Hourly & daily	SODA	Air pressure
	Dayl	Hour	Daily	SODA	Length of the day from sunrise to sunset
	Nsun	Hour	Daily	IRIMO (IRIMO, 2018)	Number of sunny hours
	C	Okta	Each 3-hours	IRIMO	Cloudiness
	Dust	Nanometer	Each 3-hours	SODA	Dust @ 550 nm
	Dummy	-	-	-	first half of the year = 1; second half of the

[4] Two data collections for temperature are acquired from IRIMO and SODA website.
[5] Hectopascal (100 x 1 pascal) - pressure units.

					year = 0.
Output	Target	kWh	Hourly & daily	IGMC (IGMC, 2017)	The amount of solar power generation per unit

The second one is called the hidden layer which is the center of all computations, weights, biases, activation functions, and calculation nodes located in this layer. Finally, the last layer is called the output layer which only shows the result of computations in the hidden layer. A brief diagram of ANN with input and output parameters is shown in Figure 2.

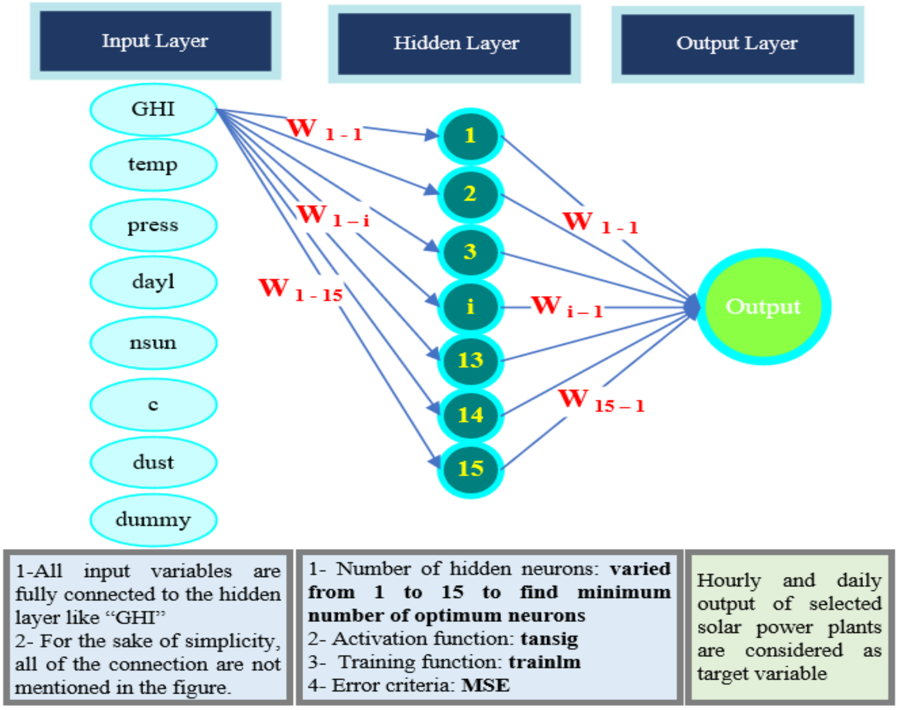


Figure 2. Schematic structure of neural networks

The purpose of the current study is to propose how to select input variables to estimate an accurate model for predicting the output of solar power plants with meteorological station data. The process of estimation of models in this paper is shown in Figure 3. After selecting the best estimation model, input variables are ranked based on the connection weight method. The proposed model can be used to select input variables to estimate the output of solar power plants. Indeed, the current study is done with real meteorological station data and it can be useful for medium to small power plants that do not have access to satellite meteorological data.

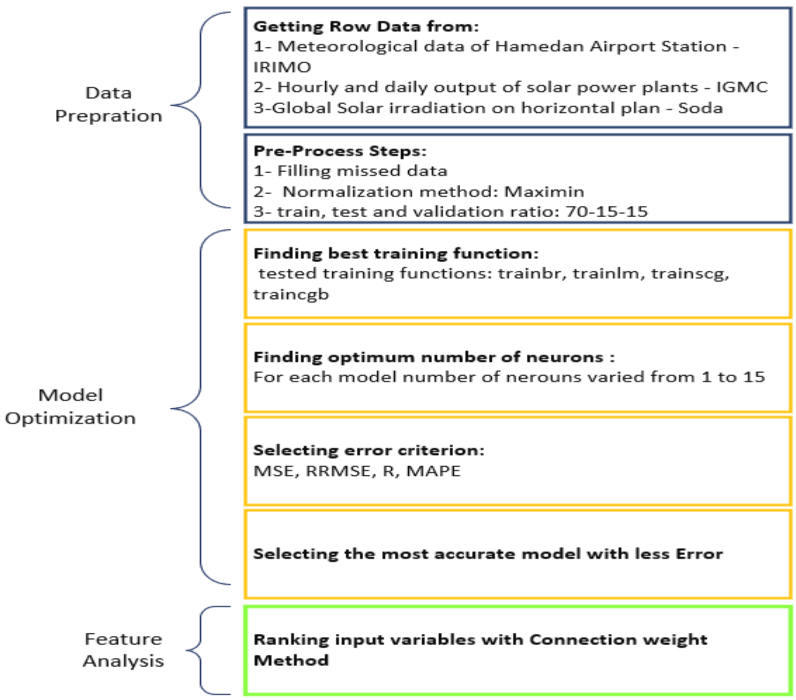


Figure 3. The process of modeling

3.1. Pre-processing steps

The algorithm of the proposed approach is given in Algorithm 1. For minimizing the estimation error, some steps are necessary for cleaning data before estimation. Preprocessing eliminates the negative effect of different input variable scales. The min-max method is used in the current study to normalize input variables and convert all input and target variables to ranges 0 and 1. In this method, the data is normalized based on the following formula:

$$x'_{ij} = \frac{x_{ij} - x_{mini\,j}}{x_{maxi\,j} - x_{mini\,j}} (x'_{maxi\,j} - x'_{mini\,j}) + x'_{mini\,j}$$

where $x_{maxi\,j} = \max x_{ij}$ indicates the maximum value of variable j and $x_{mini\,j} = \min x_{ij}$ indicates the minimum value of variable j. Moreover, $x'_{maxi\,j}$ and $x'_{mini\,j}$ are maximum and minimum values of variable j, respectively, in the new range. In the current study, $x'_{maxi\,j}$ and $x'_{mini\,j}$ are considered 1 and 0.1 to normalize data between (0.1,1).

3.2. Select the best training function

To choose the best training function, it is necessary to compare the performance of different training functions. Daily data from the Khalij-Fars power plant were used for this purpose. Each neuron is trained 30 times with a specific training function. As the weights and biases are assigned random values at every execution of the neural network, the error is sensitive to sampling and it was not satisfactory to rely on the results of one run of the model. Therefore, for each neuron i, the network is trained 30 times with each of these training functions. The function with the least mean squares error was selected as the optimum training function.

3.3. Find the optimal number of neurons

Data were randomly divided into three sets of training, validation, and test data with a ratio of 15-15-70. In each model, for one neuron up to 15 neurons, the model was repeated 100 times so that we could find the minimum optimal neuron. The average MSE error of 100 repetitions was calculated for each neuron. The model that was minimized in terms of the average MSE error of the validation dataset was selected as the optimal network.

3.4. Estimate model and rank input variables

In the present study, the logsig transfer function is used for the Multi-Layer Perceptron neural network (MLP). In the following, we turn to the selection of the optimal neural network model. The first step is to select the appropriate training function and the second step is to find the optimal number of neurons.

Algorithm1. Algorithm of the proposed approach

- **Input variables**: i = [GHI, Temp, Press, Dayl, Nsun, C, Dust, Dummy]
- **Target variable**: output of solar power plants
- **Final result**:

1-Predicting the output of solar power plants

2- Ranking the importance of each variable by the connection weight method

- **Data pre-process steps**:

Step1: Normalizing data with maximin method between (0.1,1) using the following equation:

$$x'_{ij} = \frac{x_{ij} - x_{mini\,j}}{x_{maxi\,j} - x_{mini\,j}} (x'_{maxi\,j} - x'_{mini\,j}) + x'_{mini\,j}$$

Step2: Find the best training function among Bayesian Regularization Backpropagation (trainbr), Levenberg-Marquardt (trainlm), Scaled Conjugate Gradient (trinscg), and Conjugate Gradient with Powel/Beale Restarts (traincgb).

Step3: Find the optimum number of neurons for each model

- **Model estimation**

Step4: Estimate the model with 8 input variables and evaluate the error criteria

Step5: Save the network weights:
1- Input to hidden layer weights: $W_{input(\,i\,)-hidden\ neuron(\,j\,)}$

2- Hidden layer to output layer weights: $W_{hidden\ neuron(\,j\,)-output}$

Step6: Eliminate one variable and estimate the model with 7 input variables and evaluate error criteria.

Step7: Save the network weights:
1- Input to hidden layer weights: $W_{input(\,i\,)-hidden\ neuron(\,j\,)}$

2- Hidden layer to output layer weights: $W_{hidden\ neuron(j\,)-output}$

- **Ranking parameters**

Step8: Connection weight method: The calculations of the method of connection weights are as follows:

1- Multiplication of the transpose of weights from the input layer to the hidden layer by the weights of the hidden layer to the output layer (Each row of this matrix represents an input).

$$W - \text{Connectionweightmethod}_{input(i)} = W_{input(i)-hiddenneuron(j)} * W_{hiddenneuron(j)-output}$$

2- Summarization of the numbers in each row of this matrix indicates the importance of its corresponding property.

$$\text{Total Rank of Input (i)} = \sum_{j=1}^{n} (W_{input(i)-hidden\ neuron(j)} \times W_{hidden\ neuron(j)-output})$$

3.5. Case study

The Iranian grid management company provided daily and hourly output data for all in-operation solar power plants. As shown in Figure 4, Hamedan is ranked the first in solar energy production. We selected two power plants with the most available data from the received data. In Hamadan province, the final outputs of two solar power plants, Amirkabir and Khalij-Fars, are considered since their establishmen.

The target variable for this study is the amount of electricity generated by selected solar power plants. In three formats, hourly, daily, and monthly, the Iran power grid management company has provided data related to the production of the Khalij-Fars power plant located in the city of Qahvand during the period of 2017/03/19 until 2018/03/20 and data related to Amirkabir, located in the city of Qerkhlar, from the date the power plant began operating, 2017/01/28, until 2018/03/20.

According to Google maps, the distance between Amirkabir solar power plant and Hamedan airport is 14.7 km (Map, 2023a) and the distance between Khalij-Fars solar power plant and Hamedan airport is 33.1 km (Map, 2023b).

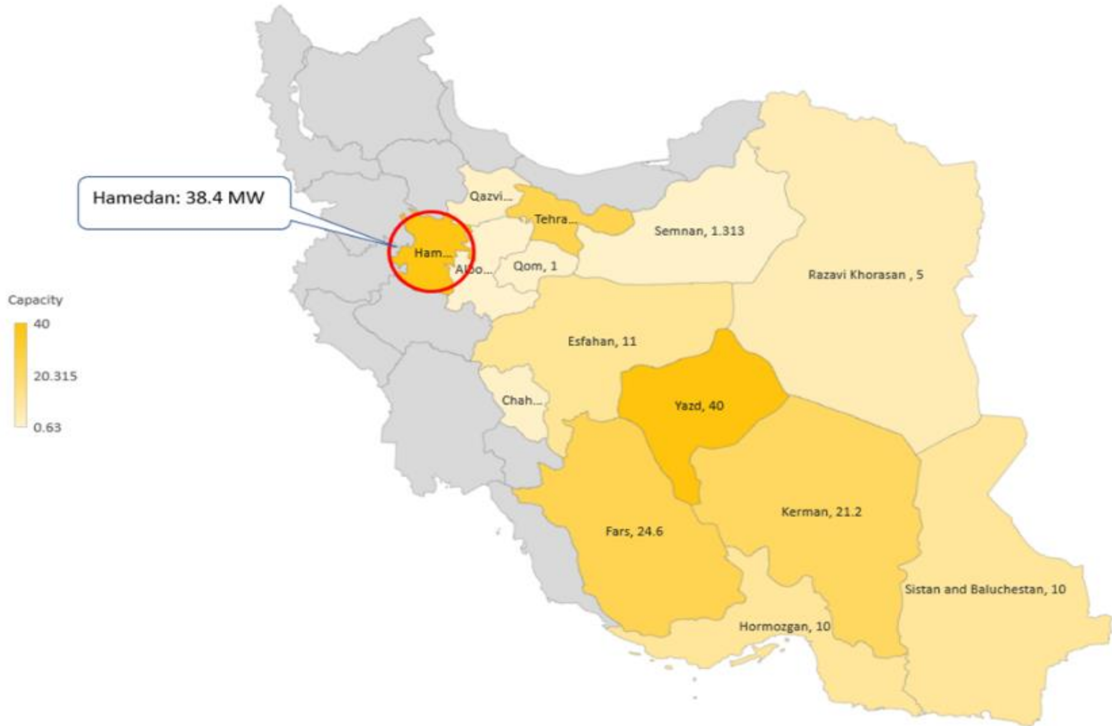


Figure 4. The total number of installed photovoltaic power plants in each province (MW) – Iran

Table 2. Description of data and the geographical location of selected solar power plants

Plant	City	Longitude	Latitude	Start date	End date	Number of daily data with zero production	Number of hourly data with non-zero production	Number of hourly data with zero production
Khalij-Fars	Qerkhlar	48.5550	34.9847	2017/03/21	2018/03/21	417	5367	10008
Amirkabir	Qahavand	48.9984	34.8586	2017/03/21	2018/03/21	426	5438	10224

In previous studies and as mentioned above, the temperature negatively affects solar collector performance (Ma et al., 2020). It is also shown in Figures 5 and 6. Although greater solar radiation occurs during warm days of the year when temperatures exceed 25 degrees Celsius, PV solar energy plants often produce the most when temperatures are below that level.

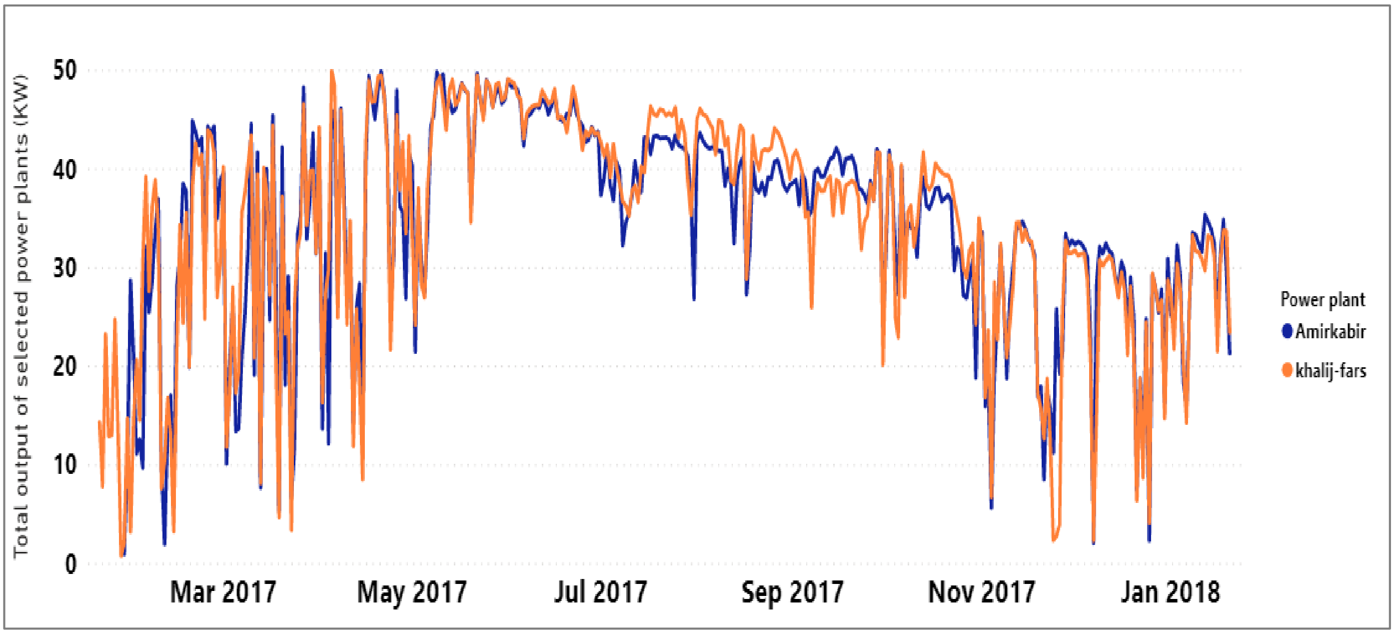
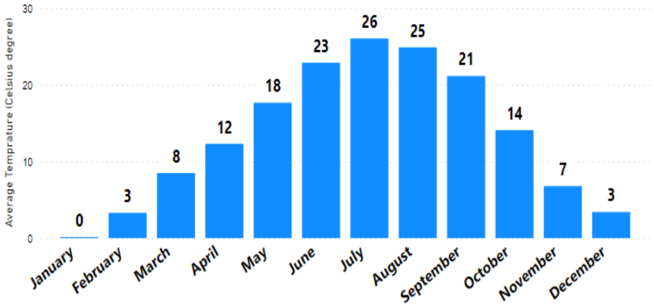
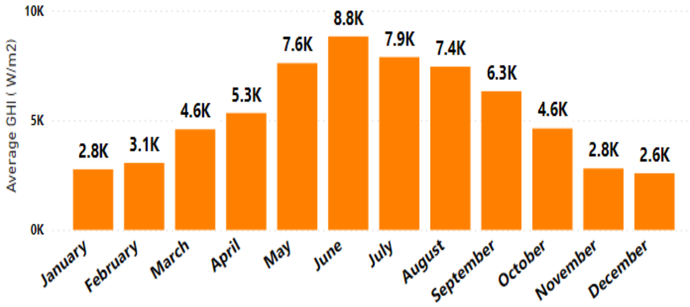


Figure 5. The daily total output of the selected solar power plants (IGMC, 2018)



6(a) Average Temperature by Month – Hamedan airport station



6(b) Average GHI by month – Hamedan airport station

Figure 6. Average Temperature & GHI by month – Hamedan airport station (IRIMO, 2018)

3.6. Assessment criteria

A prediction model's accuracy must be checked after evaluation. For the time series Y_t which contains k elements, the time series f_t is estimated. Table 3 presents the performance evaluation functions and a brief description of them.

Table 3. Assessment criteria and descriptions

The formula of the performance function	Description	Performance analysis
$\text{MAPE} = \frac{\sum_{t=1}^{k}\left\|\frac{y_t - f_t}{y_t}\right\|}{k}$	Mean absolute percentage error (MAPE) is used to differentiate between prediction models and to select the optimal model	• MAPE < 10 %: High predictive power • 10 % < MAPE < 20 %: Good predictive power • 20 % < MAPE < 50 %: Logical predictive power • MAPE > 50 %: Incorrect prediction
$\text{MSE} = \frac{\sum_{t=1}^{k}(y_t - f_t)^2}{k}$	Mean square error (MSE) is one of the most popular performance evaluation criteria. The most important drawback of this criterion is that it increases the effect of large errors.	Nearest to zero shows a more accurate forecast
$\text{RRMSE} = \frac{\sqrt{\frac{\sum_{t=1}^{k}(y_t - f_t)^2}{k}}}{\sum_{t=1}^{k} y_t} * 100$	This error is the root of RMSE and is called (RRMSE). In the present study, the model was selected with the highest accuracy by comparing the two criteria of MAPE and RRMSE.	• RRMSE < 10 %: High predictive power • 10 % < RRMSE < 20 %: Good predictive power • 20 % < RRMSE < 30 %: Logical predictive power • RRMSE > 30 %: Incorrect prediction
$R = \frac{\sum_{i=1}^{k}(f_i - \bar{f})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{k}(f_i - \bar{f})^2}\sqrt{\sum_{i=1}^{k}(y_i - \bar{y})}}$	The correlation coefficient called (R) represents the percentage of a total change of the dependent parameter, which can be explained by independent parameters. If the value of the correlation coefficient is exactly 1, it indicates that 100% of the dependent parameter changes are explained by independent parameters.	• R > 0.9 shows well-fitted models • 0.8 < R < 0.9 shows good accuracy • 0.5 < R < 0.8 shows weak fitted models

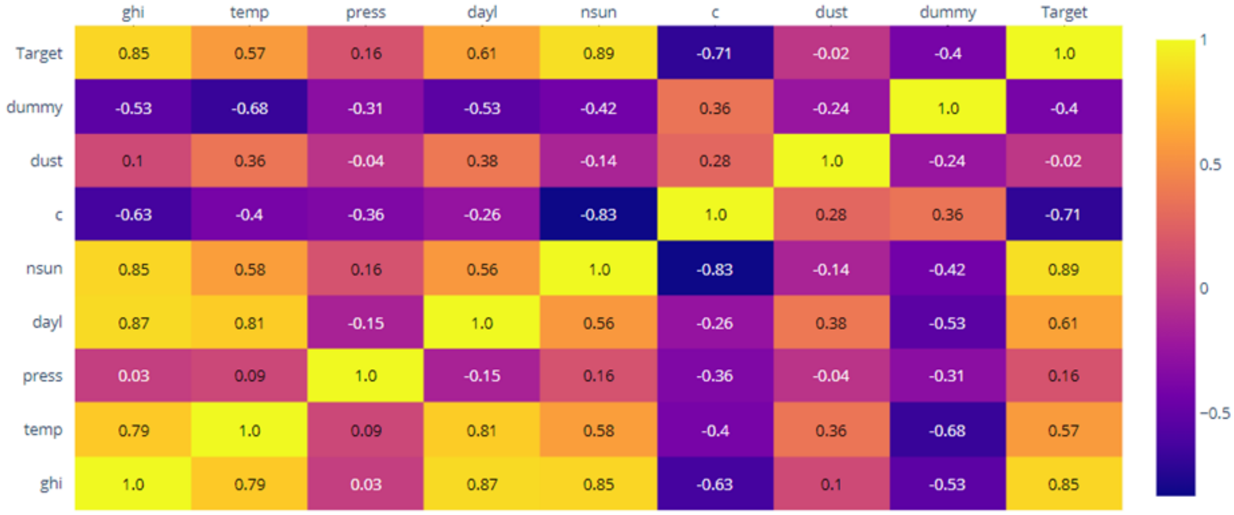
4. RESULTS AND DISCUSSION

4.1. Heatmap correlation of all variables in Khalij-Fars dataset

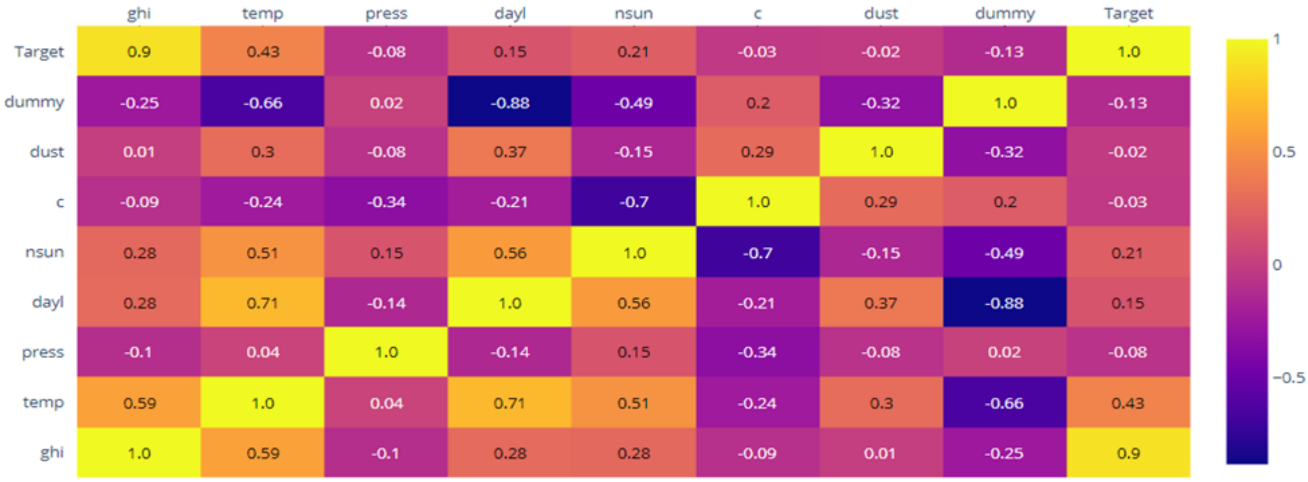
The heatmap correlation of all input and output variables for the Khalij-Fars power plant is shown in Figure 7. As is shown in Figures 7-a and 7-b, there are some differences between the correlation of parameters in daily and hourly inputs and target. In daily estimation, three more correlated factors are Nsun GHI and Dayl. In hourly estimation, 3 more correlated factors are GHI temperature and Nsun.

4.2. Heatmap correlation of all variables in the Amirkabir dataset

The heatmap correlation of all input and output variables of Amirkabir solar power plant is shown in Figure 8. Based on Figure 8-a, six input variables correlate more than 0.6, showing that our daily prediction will be more precise than the hourly forecast. In daily estimation, three more correlated factors are Nsun, GHI, and Dayl. In hourly estimation, three more correlated factors are GHI temperature and Nsun; however, as shown in Figure 8-b, just one input variable (GHI) has a correlation above 0.5 and it shows that hourly forecasts are highly dependent on the GHI data. This result shows that in both power plants, three more correlated factors in daily and hourly estimations are the same; therefore, it is confirmed that the most important factor should be selected based on the granularity of the model.

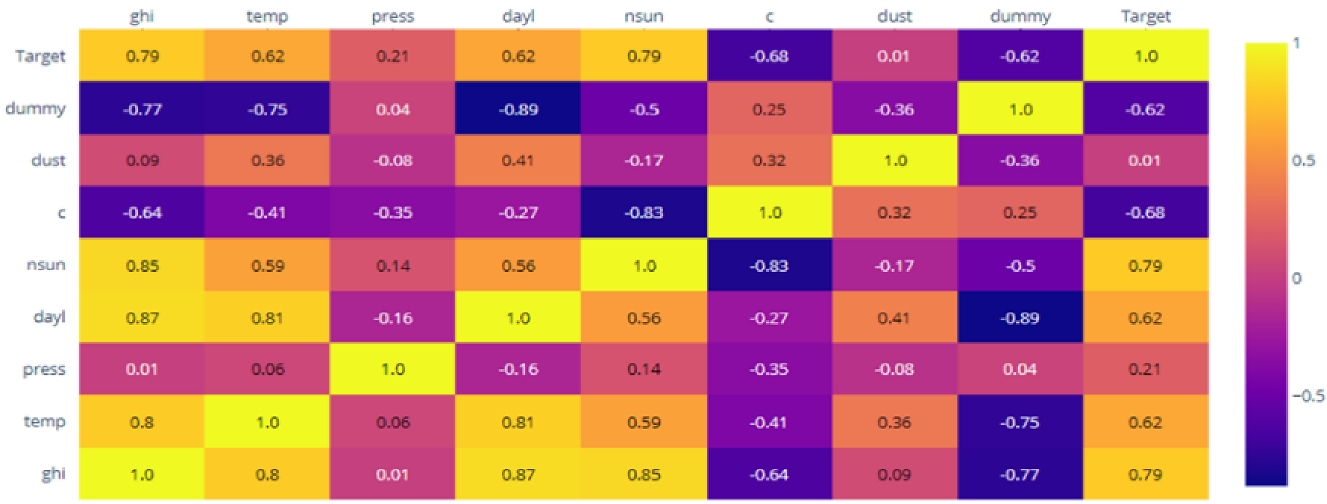


7(a) Heatmap correlation of daily inputs



7(b) Heatmap correlation of hourly inputs

Figure 7. Heatmap correlation of inputs and target - Khalij-Fars



8(a) Heatmap correlation of daily inputs

8(b) Heatmap correlation of hourly inputs

Figure 8. Heatmap correlation of inputs and target- Amirkabir

4.3. Selecting the best training function

As shown in Figure 9, the mean squared error of four training functions is compared to select the best function with minimum squared error. For each training function, the number of neurons varied from 1 to 10 to see the result of an increasing number of neurons on the performance of the train function. As shown in Figure 9, the network error with the trainlm and trainbr training functions is similarly lower than trinscg and traincgb. The model error with the trainlm training function and 7 neurons (i = 7) has the least error. As the selection of the optimal neuron and training function should be based on the selection of the neuron with the least average squared error in the validation data set, the Cumulative error distribution diagram is used to compare the four training functions.

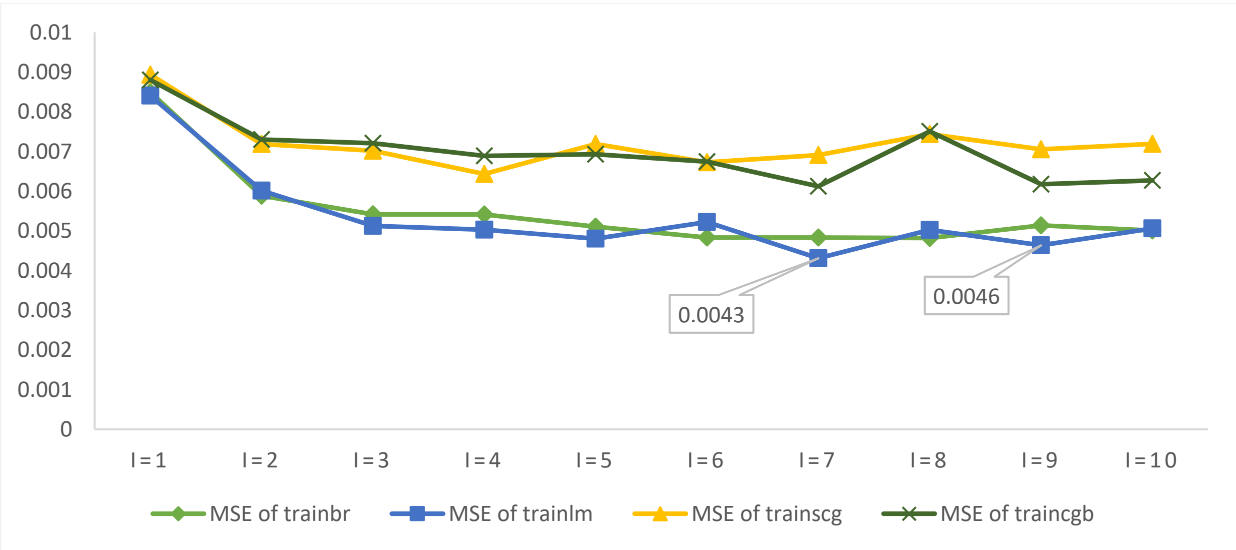


Figure 9. Selecting the best training function

4.4. Finding optimized network

The initial phase of our study will examine the network outputs of each solar power plant. First, a network with 8 parameters is estimated. Then, one of the features is removed from the model. A feature is omitted from every model, The optimal network structure (based on the least number of neurons and the best training function) is determined by the least amount of error in the validation data set. Errors are more likely to occur when a variable is highly correlated with a target variable. The optimal models are selected based on the MSE, RRMSE, R, and MAPE values. According to the correlation-weight method, the variables are then ranked based on their importance. Reported errors are calculated based on the average of 100 replicates.

4.4.1. MSE criterion

As shown in Figure 10, the hourly forecast with zero production in the Khalij-Fars power plant has the least error. If we increase the number of data, then MSE errors will decrease. Therefore, if we compare only MSE errors, we cannot determine the most accurate models between hourly forecast and daily forecast.

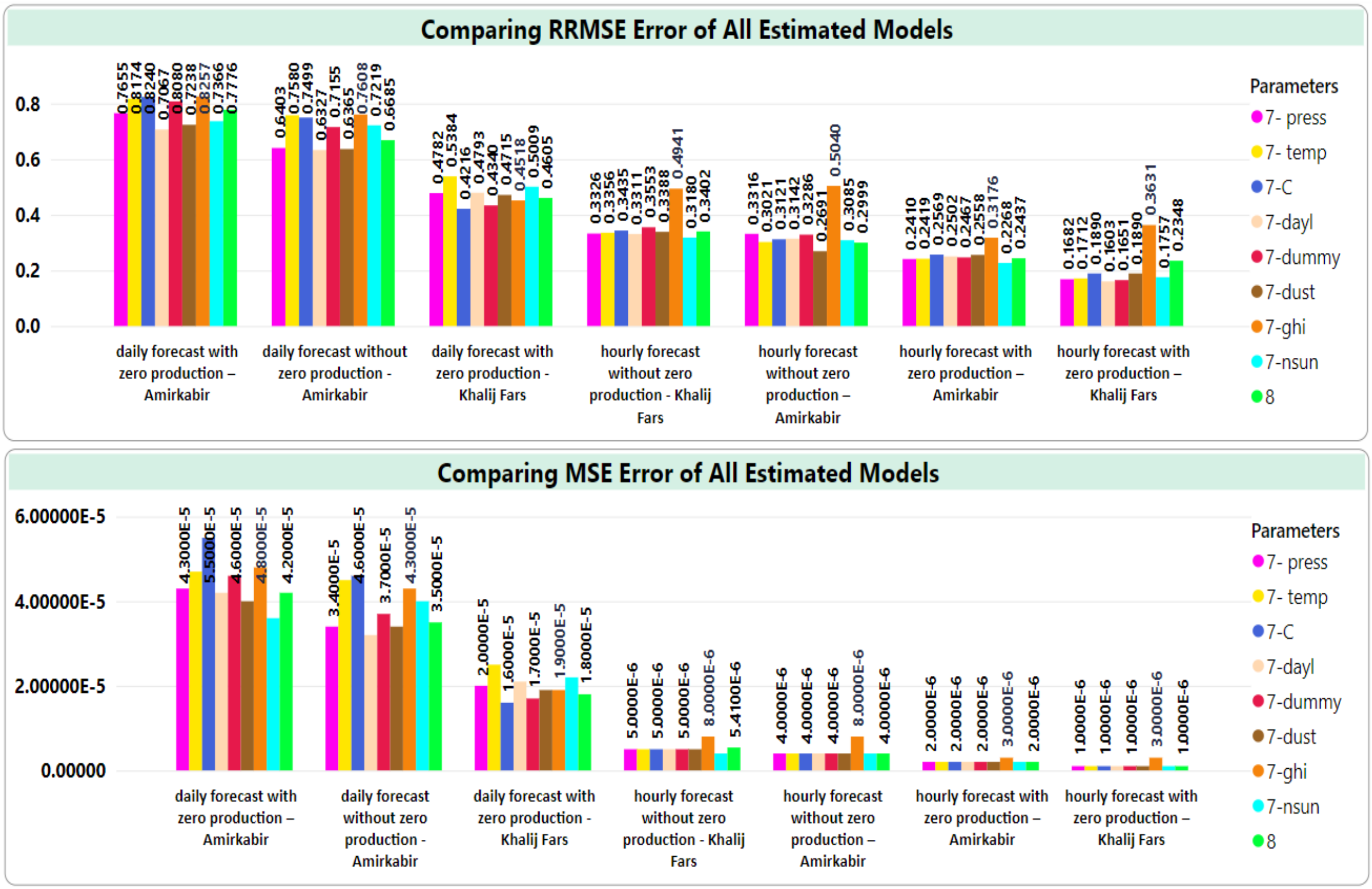


Figure 10. MSE and RRMSE error criteria of forecast models

4.4.2. RRMSE criterion

As shown in Figure 10, the hourly forecast with zero production in the Khalij-Fars power plant has the least RRMSE error**.**

4.4.3. R criterion

In Figure 11, hourly forecast with zero production in the Khalij-Fars power plant model with 7 input variables without the air pressure has the nearest value of 1. It means that the elimination of the pressure from the input variables will increase the R and decrease the MAPE and it shows that air pressure does not have a major impact on the output of solar power plants.

4.4.4. MAPE criterion

As shown in Figure 11, Khalij-Fars daily forecast with zero production without considering air pressure has the lowest MAPE error. In all the estimated models, elimination of GHI has the most negative effect on the performance of estimation and this shows the obvious high impact of GHI on the output of the solar power plant.

Compared to estimations of the Amirkabir power plant, the Khalij-Fars power plant has the least errors. To improve forecasting at Amirkabir, more accurate data is needed. In the absence of data on cloudiness parameters and the number of sunny hours, Hamadan airport data was used, which is approximately 15 kilometers away from the power plant. Note that the number of sunny hours was only reported once during the day and this variable was reported only at airport meteorological stations. It was expected that the number of sunny hours would be highly correlated with the amount of energy produced; however, the frequency of this variable did not line up with the hourly forecast (it was reported every 3 hours in the IRIMO dataset). As a result, removing this variable improves model performance in most of the hourly forecast models.

4.5. Importance analysis of input variables by connection weight method

In all the forecasting models, the connection weight method is run and the most important input parameters are ranked. The most important factor in all of the models is global horizontal irradiance, as illustrated in Figures 12. The second important parameter for the Khalij-Fars solar power plant (which is the more accurate forecast) is the temperature, and the third is the number of sunny hours. According to the connection weights method, the number of sunny hours appears to be an important variable in the model. Recent studies have supported this finding (Bugała et al., 2018; Loghmari et al., 2018; Rao et al., 2018). The number of sunny hours has been reported as an important and influential variable in the amount of solar energy production. This demonstrates the importance of measuring the number of sunny hours at meteorological stations on an hourly basis. As mentioned above, the best prediction model for the Amirkabir power plant was introduced the daily forecast mode with zero production and without air pressure.

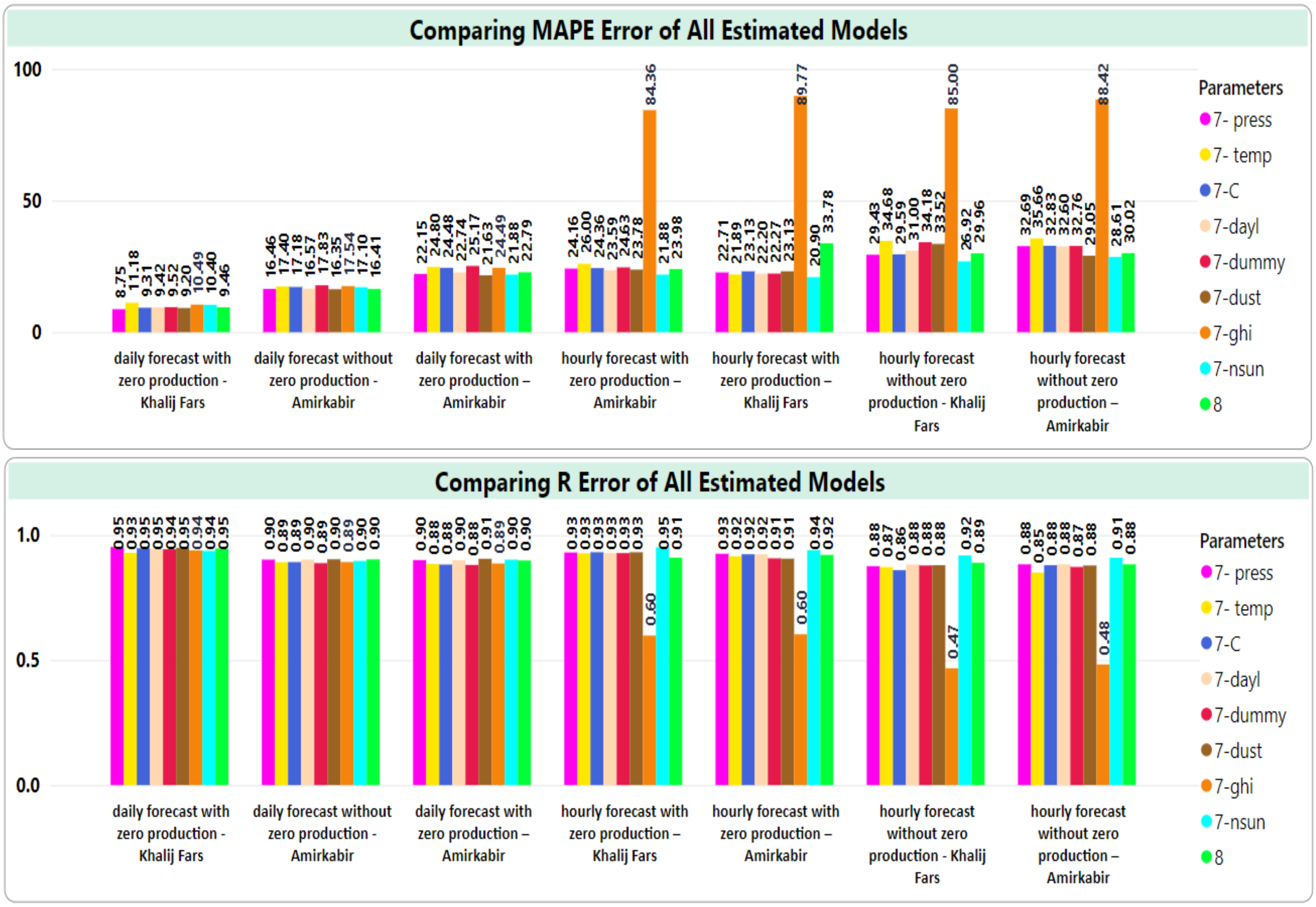


Figure 11. R and MAPE error criteria of the forecast models

Table 4. Finding minimum error in all the estimated models

Model	MSE	RRMSE	MAPE	R
Hourly forecast with zero production Khalij-Fars power plant	*	*		
Daily forecast with zero production Khalij-Fars power plant			*	*

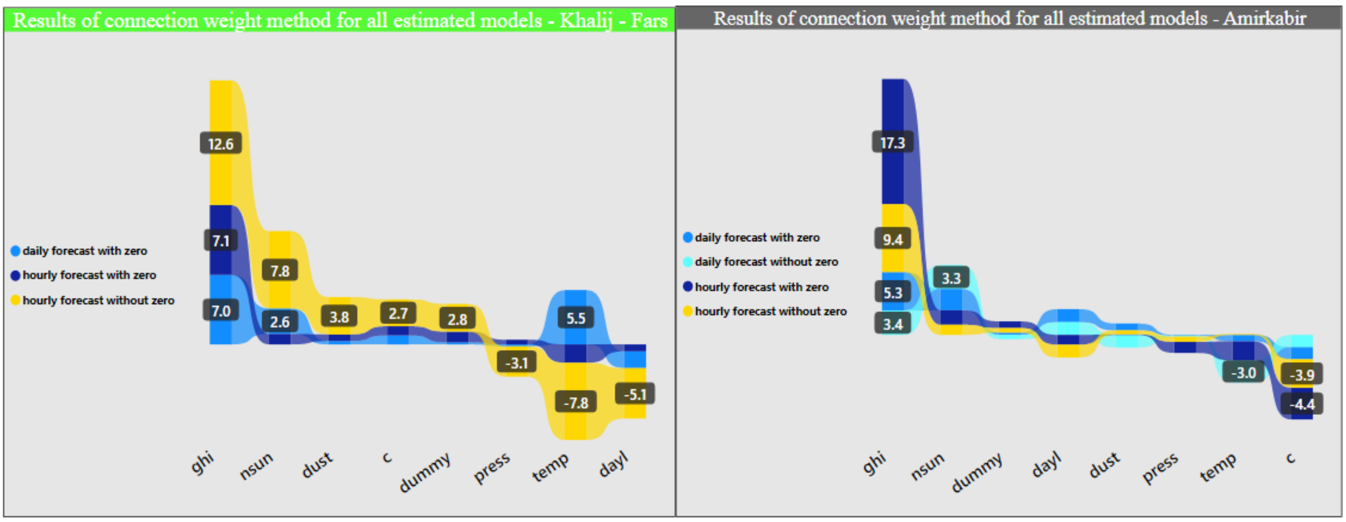


Figure 12. Result of the connection weight method

Consequently, this mode should be used to analyze the importance of input variables since it is the most accurate. In the case of daily forecast mode without zero production with seven variables without dust, the most important variable is global horizontal irradiance; the second one is the number of sunny hours; and the third one is temperature.

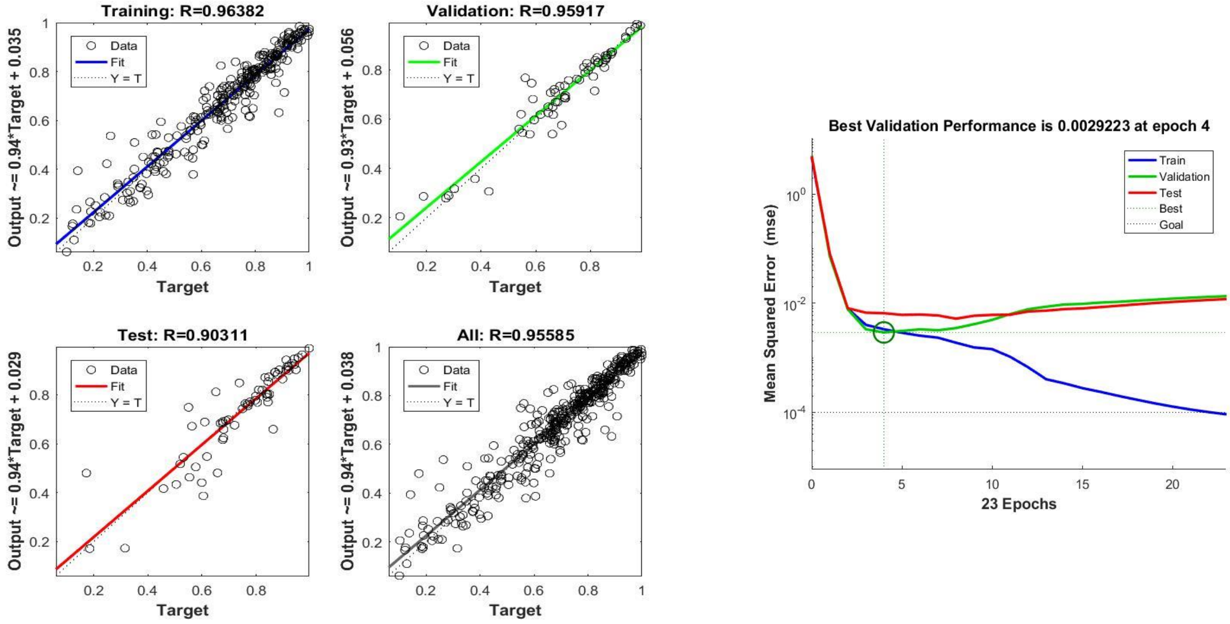
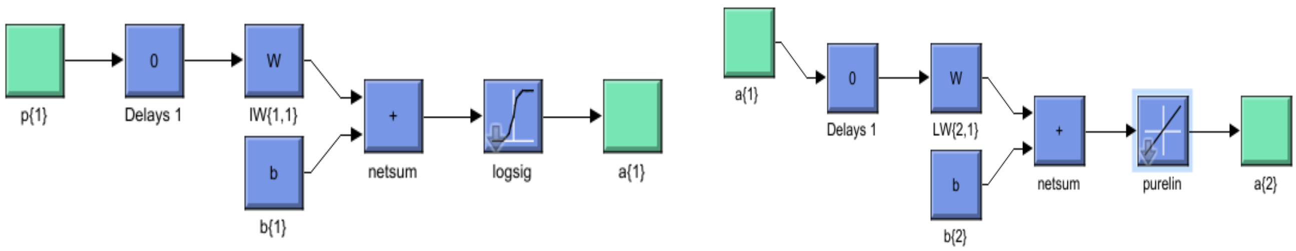


Figure 13. Performance plot of the best-estimated model Khalij-Fars daily forecast without considering air pressure



Structure of input layers to the hidden layer

Structure of the hidden layers to output

Figure 14. Structure of the estimated neural network

According to the results in Figure 13, all datasets show that the value of R is greater than 0.90; thus, the results of the connection weight method for Khalij-Fars daily forecast without considering air pressure with 6 neuron is considered the best-estimated model. Figure 14 shows the structure of the estimated network.

The equation of best-estimated model (Khalij-Fars daily forecast) without considering air pressure with 6 neurons is reported below:

$$W_{input(i)-hidden\ \ neuron(j)} * input(i) = \alpha_j$$
$$O_1 = f(\alpha_1) = f((-2.27 * ghi) + (1.74 * temp) + (2.50 * dayl) + (0.64 * nsunh) + (2.71 * c) + (-0.73 * dust) + (-1.21 * dummy) - \phi_1) = \frac{1}{1+e^{-\alpha_1}}$$
$$O_2 = f(\alpha_2) = f((1.32 * ghi) + (-0.30 * temp) + (0.01 * dayl) + (1.82 * nsunh) + (1.99 * c) + (-0.62 * dust) + (0.02 * dummy) - \phi_2) = \frac{1}{1+e^{-\alpha_2}}$$
$$O_3 = f(\alpha_3) = f((0.76 * ghi) + (2.52 * temp) + (1.11 * dayl) + (0.80 * nsunh) + (0.90 * c) + (-0.27 * dust) + (0.99 * dummy) - \phi_3) = \frac{1}{1+e^{-\alpha_3}}$$
$$O_4 = f(\alpha_4) = f((-1.56 * ghi) + (1.28 * temp) + (-0.34 * dayl) + (-0.17 * nsunh) + (-2.82 * c) + (1.05 * dust) + (-2.98 * dummy) - \phi_4) = \frac{1}{1+e^{-\alpha_4}}$$
$$O_5 = f(\alpha_5) = f((0.29 * ghi) + (1.71 * temp) + (-1.29 * dayl) + (0.83 * nsunh) + (-1.51 * c) + (1.57 * dust) + (-2.06 * dummy) - \phi_5) = \frac{1}{1+e^{-\alpha_5}}$$
$$O_6 = f(\alpha_6) = f((1.36 * ghi) + (-2.18 * temp) + (-1.44 * dayl) + (0.19 * nsunh) + (-1.78 * c) + (1.45 * dust) + (0.93 * dummy) - \phi_6) = \frac{1}{1+e^{-\alpha_6}}$$
$$O_{Final} = f((-0.98 * O_1) + (0.94 * O_2) + (1.20 * O_3) + (1.19 * O_4) + (-0.45 * O_5) + (0.72 * O_6)) = (-0.98 * O_1) + (0.94 * O_2) + (1.20 * O_3) + (1.19 * O_4) + (-0.45 * O_5) + (0.72 * O_6)$$

5. CONCLUSIONS

By considering meteorological factors and ranking their importance on solar power plant output, this research attempted to provide an optimal model to predict solar energy production more accurately. This study examined the effects of eight variables on solar power plant output. Due to the increase in the number of dusty days in the country and the negative effect of the dust parameter on the amount of solar energy production, the objective was to determine the most effective input variable for solar energy production using a connection weight method. Out of the eight variables studied in this paper, air pressure had the least effect on solar energy production. Solar global irradiance on the horizontal plane, air temperature, and the number of sunny hours were the most significant variables in the present study. The amount of global radiation on the horizontal plane was retrieved from the SODA website and is not currently reported by meteorological stations across the received data from IRIMO. Besides, calculating this variable using radiation data is not difficult. Air temperature is the second most important variable affecting the amount of solar energy production. The third place in the Khalij-Fars power plant was the number of sunny hours. Day length and the number of sunny hours were two other important variables that we encountered in the hourly forecasts because they were reported once a day. In total, for both power plants, four parameters of solar radiation, number of sunny hours, temperature, and cloudiness were introduced as the most effective parameters in the amount of solar energy production. Because the frequency measurement parameter was the number of hours of daily sunshine, this parameter was only effective in daily forecasting. The Meteorological Organization must make arrangements for monitoring and measuring these parameters on an hourly basis.

The country currently has solar power plants connected to the electricity distribution network in 18 provinces. The provinces of Kermanshah, Markazi, Ilam, and Lorestan are among those with good potential for establishing solar power plants, but have yet to do so. In addition to helping the country become self-sufficient in electricity generation with the development of solar power plants in temperate regions and border launches, with proper planning, surplus production can be exported to neighboring countries. In the present study, much time was wasted trying to format meteorological data correctly. Several hours were spent cleaning the data because the data was not recorded correctly at non-airport stations. Due to the lack of data on some days, the data of some variables were replaced with the data of the SODA site. Since the development of renewable power plants is a good alternative to fossil fuel power plants, it is necessary for the national meteorological organization to review the structure of the collected data and to be as sensitive as possible in recording the data. Given the location of Hamedan province and its proximity to Iraq, the study of the effect of dust storms on the reduction of solar energy production is one of the challenging issues in predicting the amount of solar power generation. It is practically unprofitable to set up a meteorological station near each solar power plant. Predicting production based on calculating the factors of radiation angle, radiation intensity, and temperature is one of the useful solutions that does not require the establishment of a meteorological station in the solar power plant and is done only based on theoretical calculations and temperature data. It shows the importance of conducting more research to develop precise prediction models in future studies.

6. ACKNOWLEDGEMENT

The authors thank the editor and the reviewers of the Journal of Renewable Energy and Environment for their very helpful suggestions and constructive comments with significantly improved quality of the manuscript. The authors would like to acknowledge Kharazmi University (Iran) for all the support provided.

REFERENCES

1. Bhavani, M., Vijaybhaskar Reddy, K., Mahesh, K., & Saravanan, S. (2021). Impact of variation of solar irradiance and temperature on the inverter output for grid connected photo voltaic (PV) system at different climate conditions. *Materials Today: Proceedings*. https://doi.org/https://doi.org/10.1016/j.matpr.2021.06.120
2. Bugała, A., Zaborowicz, M., Boniecki, P., Janczak, D., Koszela, K., Czekała, W., & Lewicki, A. (2018). Short-term forecast of generation of electric energy in photovoltaic systems. *Renewable and Sustainable Energy Reviews*, *81*, 306-312. https://doi.org/https://doi.org/10.1016/j.rser.2017.07.032
3. Carra, E., Ballestrín, J., Polo, J., Barbero, J., & Fernández-Reche, J. (2018). Atmospheric extinction levels of solar radiation at Plataforma Solar de Almería. Application to solar thermal electric plants. *Energy*, *145*, 400-407. https://doi.org/https://doi.org/10.1016/j.energy.2017.12.111
4. EIA. (2021). *Country analysis executive summary: Iran*. T. U. S. E. I. Administration. https://www.eia.gov/international/content/analysis/countries_long/Iran/pdf/iran_exe.pdf
5. Ghritlahre, H.K., & Prasad, R.K. (2018). Application of ANN technique to predict the performance of solar collector systems - A review. *Renewable and Sustainable Energy Reviews*, *84*, 75-88. https://doi.org/https://doi.org/10.1016/j.rser.2018.01.001
6. Gorjian, S., Ghobadian, B., & Tavakkoli Hashjin, T. (2015). Modeling of solar radiation potential in Iran using artificial neural networks. *Journal of Agricultural Science and Technology (JAST)*, *17*(7), 1707-1723. http://jast.modares.ac.ir/article-23-7341-en.html
7. Halabi, L.M., Mekhilef, S., & Hossain, M. (2018). Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. *Applied Energy*, *213*, 247-261. https://doi.org/https://doi.org/10.1016/j.apenergy.2018.01.035
8. IGMC. (2017). *Statistics and information request system*. Retrieved 2018-07-01 from https://www.igmc.ir/statistical-center
9. IGMC. (2018, 2018/07/01). *Statistical center - IGMC*. https://www.igmc.ir/statistical-center
10. IRIMO. (2018). *I.R. Of IRAN Meteorological Organization*. Retrieved 2018/09/01 from https://data.irimo.ir/
11. Jahani, B., Dinpashoh, Y., & Raisi Nafchi, A. (2017). Evaluation and development of empirical models for estimating daily solar radiation. *Renewable and Sustainable Energy Reviews*, *73*, 878-891. https://doi.org/https://doi.org/10.1016/j.rser.2017.01.124
12. Jung, Y., Jung, J., Kim, B., & Han, S. (2020). Long short-term memory recurrent neural network for modeling temporal patterns in long-term power forecasting for solar PV facilities: Case study of South Korea. *Journal of Cleaner Production*, *250*, 119476. https://doi.org/https://doi.org/10.1016/j.jclepro.2019.119476
13. Khosravi, A., Koury, R.N.N., Machado, L., & Pabon, J.J.G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. *Journal of Cleaner Production*, *176*, 63-75. https://doi.org/https://doi.org/10.1016/j.jclepro.2017.12.065
14. Loghmari, I., Timoumi, Y., & Messadi, A. (2018). Performance comparison of two global solar radiation models for spatial interpolation purposes. *Renewable and Sustainable Energy Reviews*, *82*, 837-844. https://doi.org/https://doi.org/10.1016/j.rser.2017.09.092
15. Ma, T., Li, M., & Kazemian, A. (2020). Photovoltaic thermal module and solar thermal collector connected in series to produce electricity and high-grade heat simultaneously. *Applied Energy*, *261*, 114380. https://doi.org/https://doi.org/10.1016/j.apenergy.2019.114380

16. Mamipour, S., Beheshtipour, H., Feshari, M., & Amiri, H. (2019). Factors influencing carbon dioxide emissions in Iran's provinces with emphasis on spatial linkages. *Environmental Science and Pollution Research*, *26*(18), 18365-18378. https://doi.org/https://doi.org/10.1007/s11356-019-05192-0

17. Map, G. (2023a). *Distance of Hamedan airport to Amirkabir solar power plant*. https://www.google.com/maps/dir/Hamadan+Airport,+Hamadan+Province,+Hamedan,+Sardar+Hamedani+Square,+Iran/Amir+Kabir+Solar+Power+Plant,+Hamadan+Province,+Iran/@34.9217361,48.4903413,12z/data=!3m1!4b1!4m14!4m13!1m5!1m1!1s0x3ff19279257e16c5:0x5fb78c39f38ea2dc!2m2!1d48.5378564!2d34.8673!1m5!1m1!1s0x3ff18fe4d74317c7:0xdaf5af4141bb3d0e!2m2!1d48.5660117!2d34.9669339!3e2

18. Map, G. (2023b). *Distance of Hamedan airport to Khalij Fars solar power plant*. https://www.google.com/maps/dir/Hamadan+Airport,+Hamadan+Province,+Hamedan,+Sardar+Hamedani+Square,+Iran/Persian+Gulf+Solar+Power+Plant,+Hamadan+Province/@34.9586888,48.5793117,11z/data=!4m14!4m13!1m5!1m1!1s0x3ff19279257e16c5:0x5fb78c39f38ea2dc!2m2!1d48.5378564!2d34.8673!1m5!1m1!1s0x3ff227025e5464e9:0x3af582d0f3d87ab2!2m2!1d48.7579169!2d35.0126439!3e0).

19. NASA. (2009). Retrieved 23-06-2019 from https://earthobservatory.nasa.gov/features/EnergyBalance/page4.php

20. Olden, J.D., Joy, M.K., & Death, R.G. (2004). An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling*, *178*(3), 389-397. https://doi.org/https://doi.org/10.1016/j.ecolmodel.2004.03.013

21. Piri, J., & Kisi, O. (2015). Modelling solar radiation reached to the Earth using ANFIS, NN-ARX, and empirical models (Case studies: Zahedan and Bojnurd stations). *Journal of Atmospheric and Solar-Terrestrial Physics*, *123*, 39-47. https://doi.org/https://doi.org/10.1016/j.jastp.2014.12.006

22. Professionals, S.E.S.F. (2021). *MERRA-2 METEOROLOGICAL RE-ANALYSIS*. Retrieved 05/17/2018 from http://www.soda-pro.com/web-services/meteo-data/merra

23. Siva Krishna Rao K., D.V., Premalatha, M., & Naveen, C. (2018). Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study. *Renewable and Sustainable Energy Reviews*, *91*, 248-258. https://doi.org/https://doi.org/10.1016/j.rser.2018.03.096

24. SATBA. (2021). *Renewable power plants geographical map*. Retrieved 08/05/2021 from http://www.satba.gov.ir/suna_content/media/image/2020/03/8144_orig.png

25. Shireen, T., Shao, C., Wang, H., Li, J., Zhang, X., & Li, M. (2018). Iterative multi-task learning for time-series modeling of solar panel PV outputs. *Applied Energy*, *212*, 654-662. https://doi.org/https://doi.org/10.1016/j.apenergy.2017.12.058

26. Vaka, A.S., & Talukdar, P. (2020). Novel inverse heat transfer technique for estimation of properties and location-specific process parameters of roof-mounted solar PV plants. *Thermal Science and Engineering Progress*, *19*, 100657. https://doi.org/https://doi.org/10.1016/j.tsep.2020.100657

27. Zaaoumi, A., Bah, A., Ciocan, M., Sebastian, P., Balan, M. C., Mechaqrane, A., & Alaoui, M. (2021). Estimation of the energy production of a parabolic trough solar thermal power plant using analytical and artificial neural networks models. *Renewable Energy*, *170*, 620-638. https://doi.org/https://doi.org/10.1016/j.renene.2021.01.129

28. Zhao, W., Zhang, H., Zheng, J., Dai, Y., Huang, L., Shang, W., & Liang, Y. (2021). A point prediction method based automatic machine learning for day-ahead power output of multi-region photovoltaic plants. *Energy*, *223*, 120026. https://doi.org/https://doi.org/10.1016/j.energy.2021.120026