Energy Reports 10 (2023) 2837–2849

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

Predictive modeling of PV solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression

Gökhan Sahin [a,*], Gültekin Isik [b], Wilfried G.J.H.M. van Sark [a]

[a] Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8A, 3584 CB, Utrecht, The Netherlands
[b] Computer Engineering Department, Igdir University, Igdir 76000, Turkey

ARTICLE INFO

Keywords:
Artificial neural network
Multiple linear regression
Solar power plant efficiency forecasting
Environmental variables
Principal Component Analysis

ABSTRACT

This study investigates the surface parameters and environmental factors affecting the energy production of a 500 kWp photovoltaic (PV) solar power plant in Igdir province. Using both the PV panel characteristics and the weather conditions specific to the power plant location, a total of 7 detailed features were included. The estimation of the power plant efficiency, a novel contribution not found in previous studies, is also a major focus. The performance evaluation of different models, including feed-forward neural networks and multiple linear regression, demonstrates the effectiveness of artificial neural networks in capturing the complex relationships between features and efficiency despite limited data availability. Principal Component Analysis (PCA) was used to reduce feature dimensions, showing that even with a reduced feature set, accurate efficiency prediction is still achievable. Prediction using PCA is one of the novelties of the paper. The effects of solar irradiation, module power, and module temperature on power plant efficiency are revealed. The results provide valuable insights for optimizing energy investments in the Igdir region and highlight the potential of artificial neural networks in energy forecasting, demonstrating their suitability for capturing complex patterns in solar power plant efficiency.

1. Introduction

Energy is a fundamental requirement for human life, and its consumption increases with the development of countries (Nadimi and Tokimatsu, 2018). However, the predominant reliance on fossil fuels, hydroelectric, and nuclear power plants to meet energy demands has resulted in significant environmental challenges. Fossil fuels contribute to global warming through carbon emissions, hydroelectric power plants cause ecological damage and droughts, and nuclear power plants pose threats to the environment and human health. Consequently, the importance of renewable and environmentally friendly green technologies has become evident due to environmental concerns, the rising energy needs, fluctuating fuel prices, increasing costs, and the risk of fossil fuel depletion (Ikram et al., 2021; Nazir et al., 2019).

Among renewable energy sources, solar energy stands out as a well-known clean energy option (Hosseini and Wahid, 2016). Solar power production does not release harmful gases, and it offers the advantage of low maintenance and repair costs. Furthermore, solar power plants are supported by initiatives such as the Kyoto Protocol and various green energy-related laws, including tariff and premium models (Karatayev et al., 2021; Bersalli et al., 2020). However, despite its efficiency, the widespread adoption of solar energy has been hindered. The high initial setup costs, dependence on weather conditions, hourly variations, and drop rates at dawn and dusk contribute to random variability and impact the cost-effectiveness and capacity of solar applications (Eltamaly and Mohamed, 2018).

The selection of an appropriate location for solar power plant establishment plays a crucial role in addressing these challenges. Estimating the impact of environmental conditions associated with a chosen location on energy efficiency becomes essential (Skiba et al., 2021; Evangelista et al., 2020). Additionally, with the emergence of "smart grid" concepts, planning and operating power systems have increasingly emphasized the deployment of renewable energy sources to achieve more reliable, efficient, and environmentally friendly systems (Zame et al., 2018; Bhattarai et al., 2023). However, due to the uncertain nature of renewable energy sources, advanced techniques are necessary for establishing and controlling optimal smart grids (Khalil et al., 2021).

In this context, the present study aims to estimate the power output

* Corresponding author.
E-mail address: g.sahin@uu.nl (G. Sahin).

https://doi.org/10.1016/j.egyr.2023.09.097
Received 5 May 2023; Received in revised form 14 September 2023; Accepted 16 September 2023
Available online 26 September 2023

of an exemplary photovoltaic (PV) system and assess the influence of various environmental factors on panel efficiency. Artificial neural networks (ANNs), a popular machine learning method, are employed to estimate the power output. ANN models have shown superior performance compared to linear regression models, as indicated by higher R^2 values and accuracy rates. While machine learning applications in energy prediction remain limited (Liu et al., 2019), recent studies have demonstrated their potential. For instance, neural networks and fuzzy logic have been used to predict solar irradiation in grid-connected solar PV plants (Shuvho et al., 2019), while support vector regression and feed-forward neural networks have been employed for short-term solar PV power plant forecasting (Fentis et al., 2017; Rana et al., 2016). Sharifzadeh et al. (2019) has conducted an extensive review of machine learning methods in renewable energy modeling and found neural networks as the best model. Li et al. (2019) estimated the PV output power using SVM. Verma et al. (2016), Varanasi and Tripathi (2019), Yadav and Chandel (2017); and Heydari et al. (2019) used various regression methods with annual data in predicting solar power plants. Gensler et al. (2016) made predictions with LSTM neural networks from deep learning algorithms in the same area and obtained good results compared to other techniques. Wang et al. (2019) and Sharma et al. (2011) analyzed weather variables in solar power prediction. More recently, AlShafeey and Csáki (2021) investigate the predictive capabilities of multiple regression (MR) and artificial neural network (ANN) models for photovoltaic (PV) energy generation, utilizing various input methods. By analyzing three years of PV power and weather data, the research demonstrates that ANN models consistently outperform MR models, with the hybrid input method showing the highest prediction accuracy. Additionally, the study emphasizes the adverse influence of poor data quality on forecasting accuracy, particularly in the structural approach.

In our investigation, we incorporated seven key variables into our analysis to comprehensively assess the performance of photovoltaic (PV) solar power plants. These variables were thoughtfully selected based on a rigorous review of pertinent literature within the field of PV solar power plant performance modeling. The chosen variables encompass five environmental factors, namely *solar irradiation*, *air temperature*, *wind speed*, *relative humidity*, and *air pressure*, which play pivotal roles in influencing the energy output of PV systems (Ghimire et al., 2019). Additionally, we included two module-related features, module power, and module temperature (Hwang et al., 2021). The selection of these seven variables was guided by their well-established recognition as influential factors in PV power generation within the scholarly domain. Detailed explanations of these variables, including their definitions and significance, can be found in Section 2.

By examining the power estimation capabilities of artificial neural networks and multiple linear regression, this study aims to contribute to the development of accurate and efficient methods for evaluating the performance of PV solar power plants under varying weather conditions and environmental factors. The objective is to forecast PV solar power efficiency in Igdir province under different weather conditions and provide valuable insights into the feasibility assessment and optimization of solar power plant projects (Fig. 1 and Table 1). In this study, MATLAB, Keras, and SPSS software were utilized for the experiments, with Keras and MATLAB employed for ANN modeling and SPSS for linear regression analysis.

Our approach highlights the following key aspects:

- Inclusion of both PV panel characteristics and specific weather conditions as detailed features for power plant efficiency estimation.
- Novel contribution in estimating power plant efficiency, not previously explored in similar studies.
- Utilization of Principal Component Analysis (PCA) to reduce feature dimensions while maintaining accurate efficiency prediction.
- Introduction of prediction using PCA as a novelty in the field of power plant efficiency analysis.

2. Methodology

In this section, we shall initiate by presenting the definition of variables followed by elucidating the data collection process. Subsequently, we will explicate the concept of artificial neural networks and multiple linear regression, providing a succinct overview of their underlying mechanisms. Furthermore, we will furnish a detailed exposition on the architecture and configuration of the artificial neural network model utilized in this study. Additionally, we will discuss the methodology employed for training and validating the model, encompassing pertinent

Fig. 1. 500 kWp grid-connected solar power plant at Igdir University.

Table 1
PV panel technical specifications. R_s: Series Resistance, R_{sh}: Shunt Resistance, Isc: Short Circuit Current, V_{oc}: Open Circuit Voltage, V_{max}: Voltage at Maximum Power, I_{max}: Current at Maximum Power, P_{max}: Maximum Power, FF: Fill Factor.

PV module			
Parameters	Specification	Parameters	Specification
Type of module	Poly-crystalline	Weight/module, kg	19.6 kg
R_s	0.385 (Ohm)	R_{sh}	5994.137 Ohm
V_{oc}	38.2285 V	I_{sc}	8.954 A
P_{max}	287.999 W	V_{max}	33.316 V
Dimension	1480 × 670 × 30 mm	I_{max}	8.644 A
FF	0.841	Efficiency	18%

aspects such as data preprocessing and the application of evaluation techniques to assess the model's performance.

2.1. Variables used for PV solar power plant efficiency

The efficiency of photovoltaic systems is the ratio of the total amount of electricity generated (e.g., in kWh per year) and the global solar radiation coming to that area (in the same unit, i.e., kWh/year) (Koc et al., 2019; Kaya et al., 2021; Sahin et al., 2020; Khandakar et al., 2019). It depends on several environmental factors (solar irradiation, air temperature, wind speed, relative humidity, and air pressure) and module features (module power, module temperature), which are described below.

Solar Irradiation (W/m^2): Solar irradiation refers to the amount of solar energy received per unit area over a given time period. It represents the intensity of sunlight reaching the photovoltaic panels and plays a crucial role in determining the energy output of the system. Higher solar irradiation levels generally result in increased electricity generation by photovoltaic systems. It is expressed in W/m^2 and typically measured with pyranometers (Koc et al., 2019; Kaya et al., 2021; Sahin et al., 2020; Notton et al., 2019; Sudirman et al., 2012).

Air Temperature (°C): Air temperature represents the degree of hotness or coldness of the surrounding air. It affects the efficiency of photovoltaic panels, as their performance is sensitive to temperature variations. Typically, photovoltaic systems experience a decrease in efficiency as the temperature rises due to the negative temperature coefficient of the panel's electrical characteristics. Energy production and efficiency increase when the temperature of the panel due to air circulation decreases in hot places that receive sunlight (Koc et al., 2019; Sahin et al., 2020; Türk et al., 2021).

Wind speed (m/s): Wind speed refers to the rate at which air molecules move in a particular direction. It influences the convective heat transfer from the photovoltaic panels, thereby affecting their operating temperature. Higher wind speeds can enhance heat dissipation, leading to lower panel temperatures and improved system performance. In order to get high efficiency from solar panels in hot climates, it is required to be mounted a few centimeters above the ground or roof, to ensure continuous air flow (from wind) and to prevent the panels from overheating. For this reason, wind speed and direction are important for solar energy efficiency (Koc et al., 2019; Türk et al., 2021)

Relative Humidity (g/m^3): Relative humidity represents the amount of moisture present in the air relative to its maximum capacity at a given temperature. While relative humidity does not have a direct impact on the electricity generation of photovoltaic systems, it can influence the soiling or accumulation of dust and other particles on the panel surfaces. Higher humidity levels may contribute to increased soiling, which can reduce the system's overall efficiency.

Air Pressure (millibars): Air pressure refers to the force exerted by the atmosphere on a unit area. While air pressure does not directly affect the efficiency of photovoltaic systems, it can influence the mechanical stress on the panels and their structural integrity. Significant changes in air pressure, such as those associated with extreme weather conditions, may impact the long-term reliability and performance of the system.

Module Power (kW): Module power represents the maximum electrical output that a photovoltaic module can deliver under standard test conditions. It is typically expressed in watts (W) or kilowatts (kW) and indicates the capacity or rating of the module. The module power directly influences the energy generation potential of the photovoltaic system (Kaya et al., 2021).

Module Temperature (°C): Module temperature refers to the temperature of the photovoltaic module itself during operation. It is influenced by factors such as solar irradiation, ambient temperature, wind speed, and heat dissipation mechanisms. Monitoring and controlling module temperature are essential for optimizing the performance and efficiency of photovoltaic systems, as higher temperatures can lead to reduced electrical output and accelerated module degradation.

2.2. Data collection process

The data utilized in this study was collected from a real photovoltaic (PV) power plant farm located in Igdir province. The data collection process involved the acquisition of PV panel features and historical weather data specific to the farm's location. The following steps were undertaken to ensure comprehensive data collection:

PV Panel Feature Data: Detailed information regarding the PV panel features was collected on a daily basis. This included parameters such as panel temperature and panel power. The data collection spanned a duration of three months to capture long-term trends.

Historical Weather Data: To analyze the impact of weather conditions on PV power generation, historical weather data for the same location was obtained. This encompassed variables such as solar irradiation, air temperature, wind speed, relative humidity, and air pressure. The weather data covered the same three-months period as the PV panel feature data.

Data Validation and Quality Assurance: Rigorous validation procedures were implemented to ensure the accuracy and reliability of the collected data. Data outliers, inconsistencies, and missing values were identified and addressed appropriately. Quality assurance measures were employed to minimize errors and uncertainties in the dataset.

Data Organization and Storage: The collected data was organized and stored in a structured manner to facilitate subsequent analysis. Proper documentation and labeling were employed to ensure the traceability and integrity of the dataset.

By following this systematic data collection process, a comprehensive and reliable dataset comprising PV panel features and historical weather data specific to the selected PV power plant farm in Igdir province was obtained. This dataset serves as the foundation for the subsequent analysis and modeling tasks conducted in this study.

2.3. Artificial neural networks

In this study, we employed feed-forward neural networks, which are a specific type of artificial neural networks (Sahin et al., 2020). Feed-forward neural networks, also known as multi-layer perceptrons (MLPs), are a fundamental type of artificial neural network widely used in machine learning and pattern recognition tasks. Feed-forward neural networks are a class of artificial neural networks where information flows in one direction, from the input layer to the output layer, without forming cycles. They are composed of multiple layers of interconnected nodes, also called neurons or units. The first layer is the input layer, which receives the initial data or features. Following the input layer, there are one or more hidden layers responsible for processing the input and extracting relevant features. The last layer is the output layer, which produces the final predictions or classifications. Neurons within a layer are connected to neurons in the subsequent layer through weighted connections, which represent the strength of the connection. Each neuron in a layer receives inputs from the previous layer, applies a non-linear activation function to produce an output, and passes it to the next layer.

Activation functions introduce non-linearity to the network, enabling it to model complex relationships in the data. Common activation functions used in feed-forward neural networks include the *sigmoid*, *hyperbolic tangent*, and *rectified linear unit (ReLU)* (Nair and Hinton, 2010). During the training process, the network learns the optimal weights and biases for the connections by minimizing a predefined loss or error function. This learning is typically performed using optimization algorithms such as gradient descent or its variants. Backpropagation (Rumelhart et al., 1986) is a widely used algorithm for computing the gradients of the loss function with respect to the network parameters, facilitating efficient weight updates. Feed-forward neural networks are universal approximators, meaning they can approximate any continuous function given enough neurons in the hidden layers. They have been successfully applied in various domains, including pattern recognition and forecasting (Ahmed et al., 2020; Wang et al., 2018). However, the performance and generalization of feed-forward neural networks heavily depend on appropriate architecture design, regularization techniques, and hyperparameter tuning.

Feed Forward Neural Networks (FFNN) map an input data (x_i) to an output (y_i) given in (x, y) pairs. Here (x_i, y_i) pairs can be specific to any problem. For example, in this study, x shows various features taken daily from the solar power plant, and y shows the calculated efficiency values based on these features. All phases of classical neural networks are shown in Fig. 2.

The Mean Squared Error (MSE) is utilized to measure the discrepancy between the predicted output generated by the neural network and the actual output (Sahin et al., 2020; Chicco et al., 2021). The aim is to minimize this error, with N representing the total number of samples.

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(actual_t - prediction_t)^2 \tag{1}$$

An optimization process is conducted to minimize the value of MSE. In this study, the Adam optimization (Kingma and Ba, 2014) algorithm is employed for this purpose.

2.4. Multiple linear regression

In the multiple linear regression method, how multiple input parameters (independent variables, predictors) affect an output parameter (dependent variable, response) is analyzed. In this study, the seven parameters collected were given as input to MLR and the efficiency parameter was tried to be estimated. The multiple linear regression formula as follows:

$$Y_i = b_0 + b_1X_1 + \cdots + b_iX_i + \cdots + b_kX_k + \varepsilon, \quad i = 1, 2, \ldots, k \tag{2}$$

In Eq. (2), Y_i is the dependent variable observed in multiple linear regression, X_i's are the independent variables, b's are the regression coefficients, ε is a fixed number. Using the least squares method, estimates of the regression coefficient in the multiple linear regression model are obtained as follows (Eq. (3)) (Maulud and Abdulazeez, 2020; Kaytez, 2020).

$$\widehat{Y}_i = \widehat{b}_0 + \widehat{b}_1X_1 + \widehat{b}_2X_2 + \cdots + \widehat{b}_kX_k + \varepsilon, \quad i = 1, 2, \ldots, k \tag{3}$$

$\widehat{Y}_i$ is the dependent variable in non-multiple linear regression, $\widehat{b}$'s are regression coefficients squared, ε is a fixed number.

2.5. Data scaling for preprocessing

The data obtained from the PV system was scaled using the following equation (Sahin et al., 2020), where Z represents the scaled values and X represents the raw values. *Mean* refers to the mean of X, and *std* refers to the standard deviation of X. Data scaling is used to normalize the range of data and bring all features to a similar scale, ensuring that no single feature dominates the learning process in machine learning models.

$$Z = \frac{X - X_{(mean)}}{X_{(std)}} \tag{4}$$

2.6. Performance metrics

In evaluating the performance of the artificial neural network (ANN)

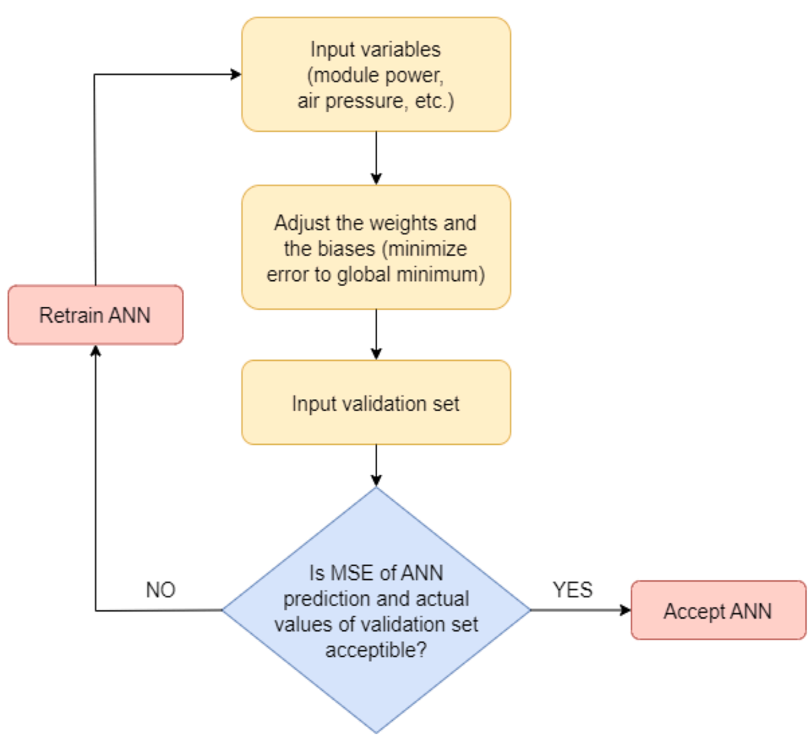


Fig. 2. The classical ANN flowchart.

and multiple linear regression (MLR) models, several metrics were utilized. The metrics used to evaluate the performance of the models are as follows:

Correlation Coefficient (r): The correlation coefficient ranges between −1 and 1, where a positive value indicates a positive linear relationship, a negative value indicates a negative linear relationship, and a value of 0 indicates no linear relationship between the variables (Eq. (5)).

$$r = \frac{(\Sigma[(x_i - \overline{x})(y_i - \overline{y})])}{\sqrt{\Sigma(x_i - \overline{x})^2}\sqrt{\Sigma(y_i - \overline{y})^2}} \tag{5}$$

Here x_i and y_i are individual data points in variables x and y, respectively. And x and $\overline{y}$ are the means of variables x and y, respectively. $\sum$ represents the summation operator, indicating the sum of values across the dataset.

Mean Absolute Error (MAE): The mean absolute error measures the average magnitude of the differences between the predicted and observed values. It provides a measure of the average absolute deviation from the actual values, disregarding the direction of the errors (Eq. (6)) (Willmott and Matsuura, 2005).

$$MAE = \left(\frac{1}{n}\right)\sum |y_i - x_i| \tag{6}$$

where y_i represents the observed or actual value, x_i represents the predicted value, and n is the total number of data points.

Root Mean Square Error (RMSE): The root mean square error measures the square root of the average of the squared differences between the predicted and observed values. It provides a measure of the average magnitude of the errors, giving more weight to larger errors due to the squaring operation. The RMSE is a commonly used metric for evaluating the accuracy of predictive models (Eq. (7)) (Willmott and Matsuura, 2005; Chicco et al., 2021).

$$RMSE = \sqrt{\left[\left(\frac{1}{n}\right)\sum (y_i - x_i)^2\right]} \tag{7}$$

Coefficient of Determination (r²): The coefficient of determination (r-squared or r^2) represents the proportion of the total variation in the dependent variable that can be explained by the independent variables in the model. It ranges from 0 to 1, where a value closer to 1 indicates a better fit of the model to the data (Eq. (8)) (Chicco et al., 2021).

$$r^2 = 1 - \left(\frac{SSres}{SStot}\right) \tag{8}$$

where $SSres$ is the sum of squares of residuals or errors and $SStot$ is the total sum of squares.

Mean Absolute Percentage Error (MAPE): The MAPE is calculated by taking the absolute difference between the actual and predicted values, dividing it by the actual value, and multiplying by 100 to express the error as a percentage. The average of these percentage errors is then computed to obtain the overall MAPE (Eq. (9)) (Chicco et al., 2021).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \widehat{y}_i}{y_i}\right| \times 100 \tag{9}$$

where n is the total number of samples, y_i represents the actual value, and $\widehat{y}_i$ represents the predicted value.

3. Experimental details

In this research, the performance of feed forward neural networks and multiple linear regression models was evaluated using data from a 500 kWp Photovoltaic (PV) Solar Power Plant. Before feeding the data into the models, a series of preprocessing steps were carried out to ensure the quality and compatibility of the dataset. These steps included data cleaning, normalization, and splitting into training and test sets using k-fold cross-validation. The dataset used in the study consisted of daily average measurements from January, February, and March 2018, obtained from the General Directorate of State Meteorology Affairs for the Igdir province. A total of 90 samples were collected daily from the solar panel, and the dataset was divided into training and test sets using k-fold cross-validation with k set to 10. This process involved dividing the data into 10-folds, with 9-folds used for training and 1-fold for testing. This separation of training and test sets was repeated 10 times to ensure consistency of the results and to reduce randomness. In the case of the FNN model, the seven input features, including solar irradiation, air temperature, wind speed, relative humidity, air pressure, module power, and module temperature, were used to predict the efficiency of the photovoltaic (PV) solar power plant. These features were normalized to ensure consistent scaling across variables. The FNN model consisted of an input layer with seven neurons, a single hidden layer with 15 neurons, and an output layer with one neuron responsible for predicting efficiency values. Based on the results, it was found that using 15 neurons in a single hidden layer yielded the best performance (Fig. 3). However, when the number of hidden layers was increased to two, the network's performance declined, indicating an issue of overfitting. The Adam optimization (Kingma and Ba, 2014) algorithm was employed with a learning rate (α) of 0.001. The neural network was trained using the Keras library (Chollet, 2015), which is implemented in Python and provides a convenient API for neural network operations, on the TensorFlow framework (Abadi et al., 2016).

Principal Component Analysis (PCA) was employed as a dimensionality reduction technique to eliminate less significant features from the data.

For the MLR model, the same seven input features were used to predict the efficiency of the PV solar power plant. The dataset was preprocessed similarly, with features normalized for consistency. Unlike the FNN, the MLR model relied on linear relationships between the input features and the target variable (efficiency). Regression analysis was performed to estimate continuous numerical values in the output layer.

In both models, the training data were used to train the models, and the test data were employed to evaluate their predictive performance. Various performance metrics, including R-squared (R^2), root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE), were used to assess and compare the performance of the FNN and MLR models.

In this study, we employed ANNs not as variable selection tools but as predictive models to assess the impact of various environmental factors and module features on the efficiency of a photovoltaic (PV) solar power plant. ANNs were utilized to estimate the power output and evaluate the influence of the selected variables on solar panel efficiency. It is essential to emphasize that ANNs were not employed for variable selection; instead, they served as advanced predictive tools capable of capturing intricate, non-linear dependencies within the data.

The choice of ANNs over traditional linear regression methods was motivated by the inherent limitations of linear models in capturing non-linear relationships among variables. Traditional linear regression relies on predefined assumptions and may not effectively represent the complex, non-linear interactions observed in the data. In contrast, ANNs have the capacity to model and predict outcomes in situations where linear methods may fall short, making them a valuable asset in addressing the research objectives.

4. Results and discussion

In this study, Neural Network Regression and Multiple Linear Regression analysis were compared. The most commonly used environmental parameters affecting solar panel efficiency were selected and examined. The objective of this study is to examine how environmental parameters impact the efficiency of panel power through the utilization

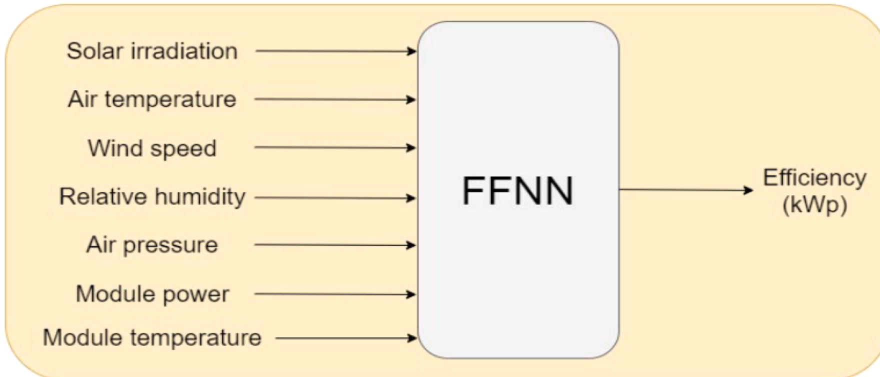


Fig. 3. Feed Forward Neural Network (FFNN) model used in the experiment.

of machine learning techniques, and the above metrics were used to compare them. Based on the findings, the artificial neural network model demonstrated an R^2 value of 0.9628, while the multiple linear regression model yielded a value of 0.901. In addition, the root of mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE) results can be seen in Table 2. Consequently, the artificial neural network model demonstrated superior performance compared to the multiple linear regression model, as it achieved better results across all evaluation criteria.

Table 3 presents the correlation coefficients between *PV solar power efficiency* and other variables. The results indicate a significant positive correlation between *PV solar power efficiency* and *solar irradiation* as well as *PV module power*. Conversely, there is a strong negative correlation between *PV solar power efficiency* and *air pressure*. These correlations are visually depicted in Fig. 4.

Fig. 5 illustrates the regression curves for the validation and test data, as well as the results for all data, showcasing the modeling performance by comparing the neural network output with the training output. The regression values for all data demonstrate a strong correlation, with values close to 1, indicating a good fit of the data to the model. The artificial neural network output closely aligns with the real data, as evidenced by the high R^2 and correlation coefficient ($r = 0.9628$) values. Additionally, the MAE and RMSE values are expected to be low, further confirming the accuracy of the model. The specific results are presented in Fig. 5.

The mean absolute error (MAE) value in the test set was found to be 25.09 as seen in Table 4. This result is within the appropriate range of 500 kWp Photovoltaic (PV) solar power plant efficiency values. In the test set, the coefficient of determination (R^2) was found to be 0.901, indicating a strong level of correlation between the predicted values and the actual values. The root mean square error (RMSE) was calculated using the same method and yielded a value of 23.89, representing the average magnitude of the prediction errors. By applying Principal Component Analysis (PCA), the number of features was reduced to two dimensions. Although the performance criteria decreased after applying PCA, the results indicate that utilizing only two dimensions is sufficient for predicting the efficiency of the solar power plant (Table 4).

To provide a clear visualization of the results, Table 5 displays the predicted values for 10 randomly selected samples. The table demonstrates the close proximity between the actual and predicted values of

Table 2
Performance criteria for artificial neural network and multiple linear regression models.

Artificial neural network				Multiple linear regression analysis			
R^2	RMSE	MAE	MAPE	R^2	RMSE	MAE	MAPE
0,9628	0,881	1450,03	9,215	0,901	1,105	1864,4	10,86

Table 3
The correlation coefficients between PV Module Efficiency and other variables.

Input parameters	PV solar power efficiency
Solar irradiation	+ 0.316
Module power	+ 0.308
Module temperature	+ 0.121
Air temperature	+ 0.096
Relative humidity	+ 0.064
Wind speed	+ 0.058
Air pressure	−0.052

the 500 kWp photovoltaic (PV) solar power plant efficiency.

In Fig. 6(a) and (b), the distribution of residuals is very close to the Gaussian curve, which means errors are concentrated around nearly zero. This error distribution is close to ideal state in data science experiments. According to these graphs, the error distribution of ANN is smoother than Multiple Linear Regression Model.

In Fig. 7, it is shown that, the variables such as solar irradiation, module power and module temperature have the largest effects on the estimation of efficiency. The status of these variables can be seen in the simple regression plots. Another point is that the wind speed and air pressure variables have less effect as in the simple regression plot (Fig. 7 and Table 6). Figs. 6 and 7 were generated using SPSS (IBM SPSS version 23 licensed program) analysis program.

Multiple Non-Linear Regression Model

The coefficients of the regression equation show the effect of input variables on the result in Table 7.

We have interpreted the analysis results in Table 7 and we found Eq. (8):

$$\textit{Efficiency} = -47934,12 + 21,05 \times b_1 + 7,8 \times b_2 - 55,76 \times b_3 + 50,61 \times b_4 + 20,14 \times b_5 + 5,58 \times b_6 - 1,26 \times b_7 \tag{10}$$

The correlation values for all parameters are presented in Table 8, along with the coefficients of the regression equation calculated based on the parameter estimates for binary interactions.

Based on the regression equation, the panel efficiency of all parameters and the identified binary interactions exhibited a significantly high value of 0.901. This observation is further supported by the ANOVA (Analysis of Variance) (Table 9), which indicates a high calculated F value. The F value represents the ratio of the variation between groups to the variation within groups. It is used to test the null hypothesis that there is no significant difference between the means of the groups being compared. A higher F value indicates a larger difference between the group means and suggests a greater likelihood of rejecting the null hypothesis. The specific value of F = 37.509 indicates a significant

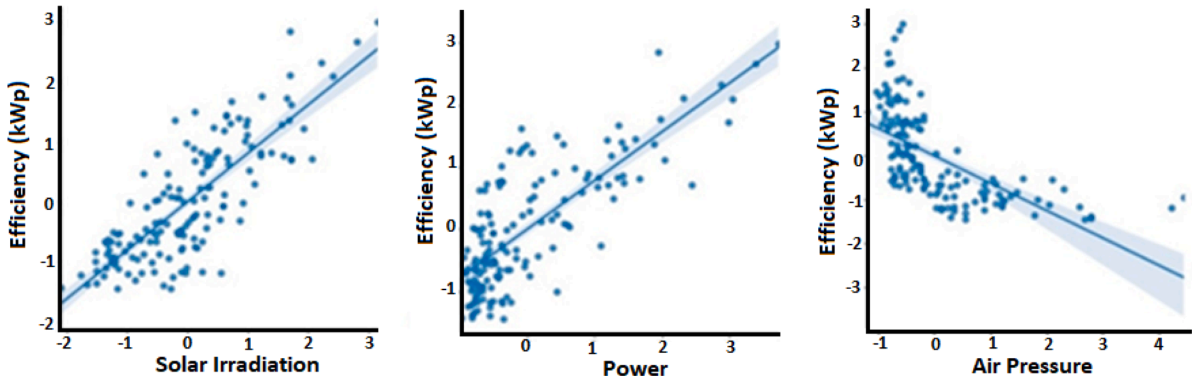


Fig. 4. Variables with strong correlations with PV solar power efficiency.

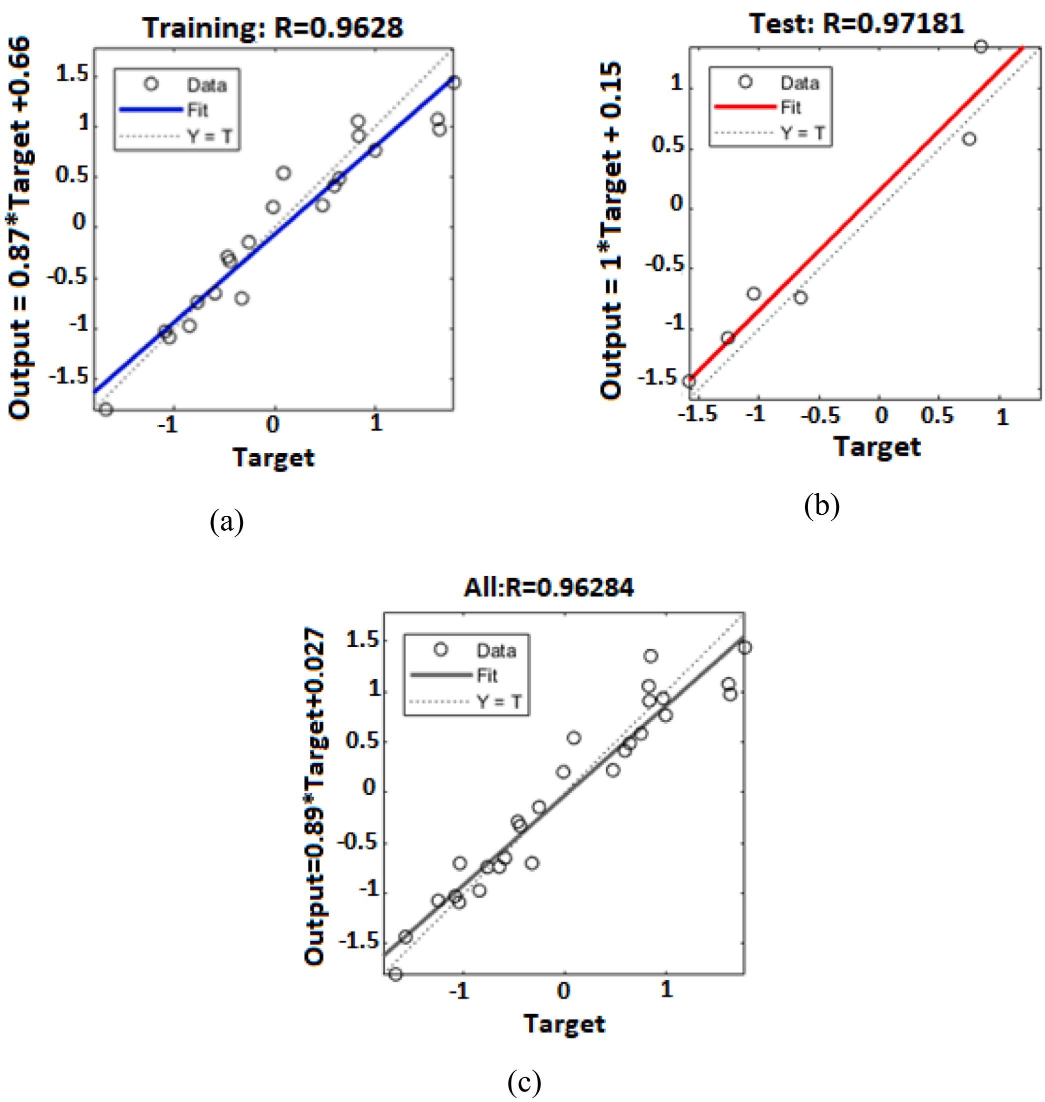


Fig. 5. Regression curve fitting of (a) training, (b) test and (c) all data for the ANN.

difference between the groups being analyzed.

$$F = \frac{15310000}{408173.343} = 37.509 \tag{11}$$

Figs. 8 to 12 depict the relationship between changing input conditions and the variation of panel efficiency. The spider web representation, displayed in column (a) of these figures, illustrates the change of the dependent variable (panel efficiency) in response to different independent variables (environmental and module factors). The focus is on observing the coherence of the variables represented in red and blue, indicating whether they change in a synchronized manner. In column (b) of the figures, the correlation between these variables is demonstrated. A strong correlation suggests that the variables move together in either a positive or negative direction.

Table 4
Experimental results with neural networks. Best values are in bold in each category.

Experiments with (#Exp. Id)		Neural network layer sizes	Correlation between true values and predictions	Testing set Mean Absolute Error (MAE)	Testing set R^2	Testing set RMSE
Normal data (7 dimensions)	#1	7–8–1	0.932	27.49	0.837	27.91
	#2	7–16–1	0.934	26.43	0.867	24.89
	#3	**7–32–1**	**0.9628**	**25.09**	**0.901**	**23.89**
	#4	7–64–1	0.937	25.39	0.871	24.09
PCA applied data (2 dimensions)	#5	**2–8–1**	**0.921**	**30.10**	**0.860**	**28.26**
	#6	2–16–1	0.908	31.60	0.802	30.86
	#7	2–32–1	0.908	31.78	0.828	31.86
	#8	2–64–1	0.912	31.29	0.845	31.55

Table 5
Randomly selected 10 samples (Solar Irradiation, Module Temperature, Wind Speed, Air Temperature, Relative Humidity, Module Power, PV Module Efficiency) and produced predictions.

Solar irradiation (W/m^2)	Module temperature (°C)	Wind speed (m/s)	Air temperature (°C)	Relative humidity (%)	Module power (W)	Solar power plant efficiency (kWh)	Predicted efficiency
721	67	4,04	2,19	43,2	493,15	1219	1219,52
612	51	6,6	5,4	85,98	449,44	1696,63	1692,61
617	52	5,07	4,66	70,84	429,21	1575,28	1585,1
672	57	3,29	3,84	69,91	467,19	1803,15	1785,2
659	34	2,35	4,2	79,12	464,48	1780,9	1793,12
658	24	1,48	5,21	80,74	494,38	1966,29	1968,36
666	36	1,49	4,45	84,02	478,31	1826,52	1810,81
654	34	2,6	2,86	79,82	464,72	1788,3	1793,2
646	35	1,88	2,9	92	456,07	1648,54	1658,39
685	34	2,08	3,94	88	463,48	1853,93	1855,68

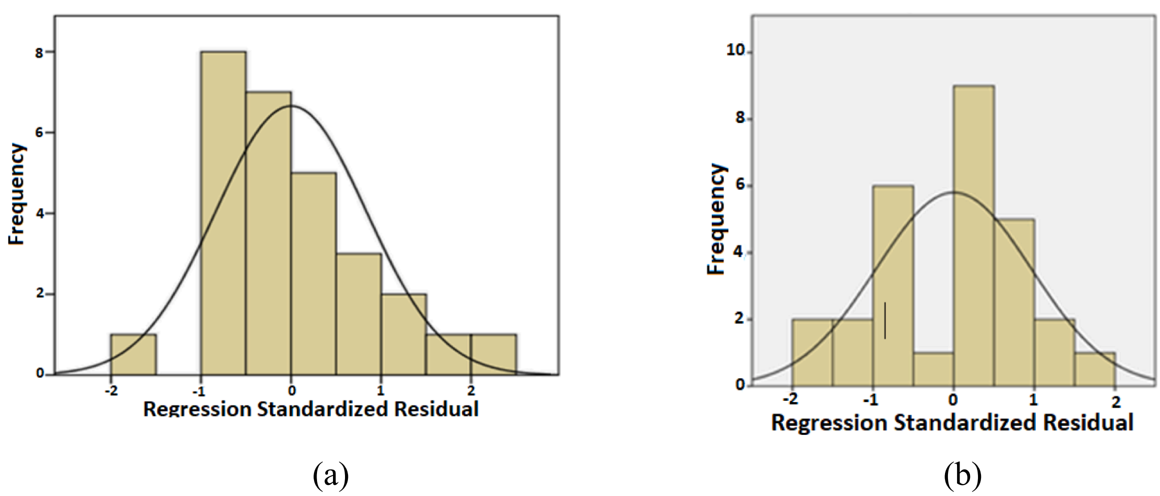


Fig. 6. (a) Regression standardized residual histogram of ANN regression, (b) Regression standardized residual histogram of multiple linear regression model.

Table 6
Independent variable importance (percentage).

	Importance	Normalized importance
Module power	,308	100,0%
Solar irradiation	,316	37,1%
Module temperature	,121	96,5%
Relative humidity	,064	20,5%
Air temperature	,096	30,8%
Air pressure	,052	16,6%
Wind speed	,058	18,5%

In Fig. 8a, the spider web representation is compared with the percentage yield and panel efficiency. The correlation between the two variables is clearly visible. Also, in Fig. 7b, the linear regression is compared with the percentage yield and panel efficiency. According to the linear regression plot, a positive strong correlation was shown between the two variables. The nova's smaller than 0.0001 ($p < 0.0001$) due to in Fig. 8b indicates a significant increase in panel efficiency. The percentage of panel activity was determined to be approximately $R^2 = 0.841$.

In Fig. 9a, the solar irradiation effects to solar efficiency, clearly. In Fig. 9b, according to the linear regression plot, a positive strong correlation was shown between the two variables. The anovas smaller than 0.0001 ($p < 0.0001$) due to in Fig. 8b indicates a significant increase in panel efficiency. The percentage of disclosure was about $R^2 = 0.651$.

In Fig. 10a, the relationship between the two variables is not clearly visible. Also, according to the linear regression plot (Fig. 10b), a positive but loose correlation was shown between the two variables. The percentage of disclosure was about $R^2 = 0.411$.

In Fig. 11a, spider web representation clearly shows the correlation between parameters. In Fig. 11b, according to the linear regression plot,

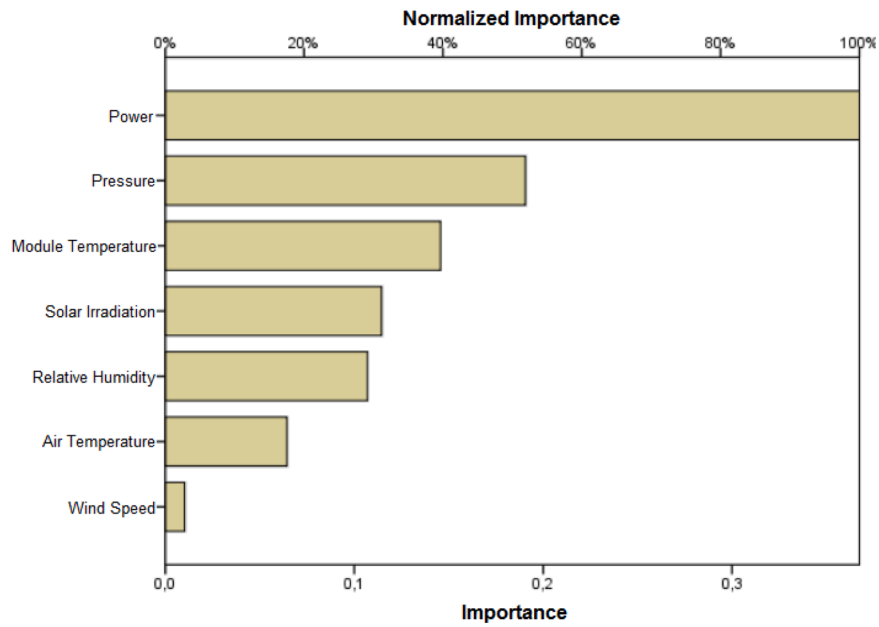


Fig. 7. Importance of input parameters for ANN regression.

Table 7
Parameter estimates of multiple nonlinear regression analysis.

Parameter estimates				
Parameter	Estimate	Std. Error	95% confidence interval	
			Lower bound	Upper bound
b0	−47 934,119	19 249,557	−88 087,992	−7780,246
b1	21,055	15,001	−10,236	52,346
b2	7,800	106,486	−214,326	229,926
b3	−55,764	141,342	−350,599	239,071
b4	50,610	20,784	7,256	93,964
b5	20,138	12,296	−5,512	45,788
b6	5,582	1,249	2,976	8,187
b7	−1,266	1,040	−3,436	,904

b_0: Bias (Intercept), b_1: Module Temperature, b_2: Wind Speed, b_3: Air Temperature, b_4: Pressure, b_5: Relative Humidity, b_6: Power, b_7: Solar Irradiation.

a positive strong correlation was shown between the two variables. The anovas smaller than 0.0001 ($p < 0.0001$) due to in Fig. 11b indicates a significant increase in panel efficiency. The percentage of disclosure of panel efficiency was determined to be about $R^2 = 0.654$.

In Fig. 12(a), spider web representation clearly shows very weak correlation between parameters. In Fig. 12b, the same weak and negative relationship is also seen in the linear regression plot. The percentage of disclosure was about $R^2 = 0.143$.

In Fig. 13(a), it is seen that there is a very weak relationship between the variables in both plots. The percentage of disclosure of panel efficiency was determined to be about $R^2 = 0.012$. The wind speed does not affect to increase or decrease of the panel efficiency.

5. Conclusion

This study investigated the surface parameters and environmental factors influencing the energy production of a 500 kWp Photovoltaic (PV) solar power plant in Igdir province. The use of feed forward neural networks and multiple linear regressions allowed for modeling of solar power plant efficiency. The results demonstrated that the system performed well despite limited data availability. Artificial neural networks have gained popularity in recent years, and in this study, they were utilized to estimate the efficiency of the solar power plant alongside multiple linear regression. The regression values indicated a successful performance of the network, with a convergence towards 1 suggesting strong regression capabilities. Performance criteria such as R^2, RMSE, MAE, and MAPE were used to evaluate the models, and the artificial

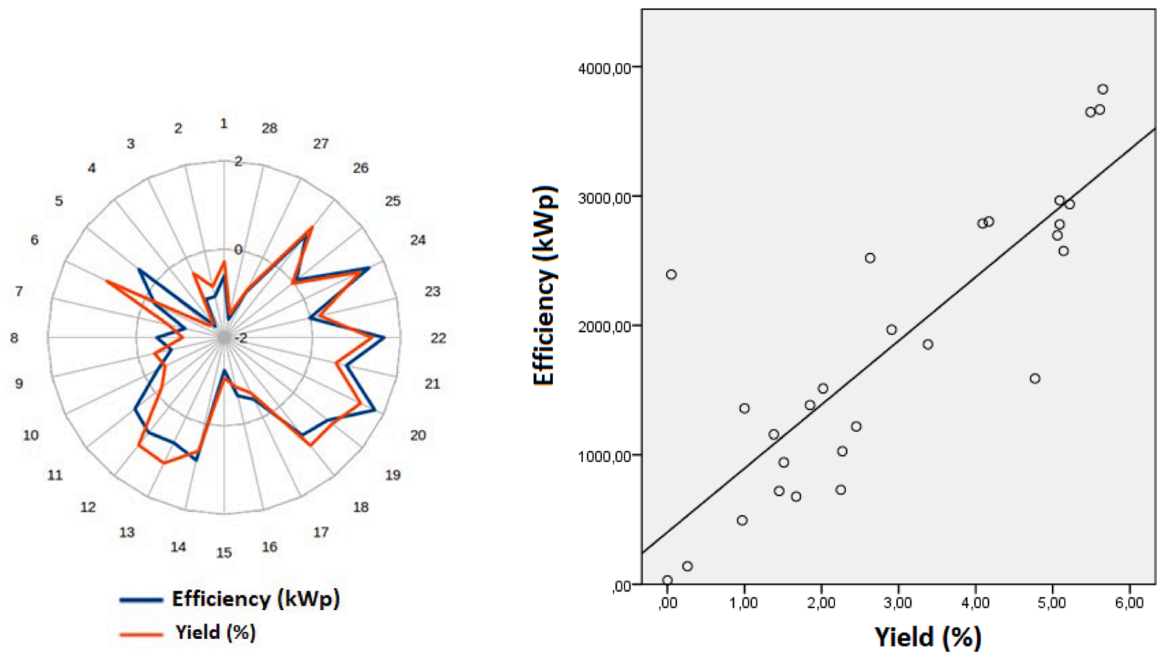
Table 9
ANOVA results for multiple linear regression analysis.

ANOVA			
Source	Sum of squares	df	Mean squares
Regression	1,225E8	8	1,531E7
Residual	8 163 466,859	20	408 173,343

Dependent variable: Efficiency
R squared = 1 - (Residual Sum of Squares)/(Corrected Sum of Squares) = 0,901 ($R^2 = 0{,}901$). *df: Degree of freedom, R^2: Adjusted coefficient of determination.*

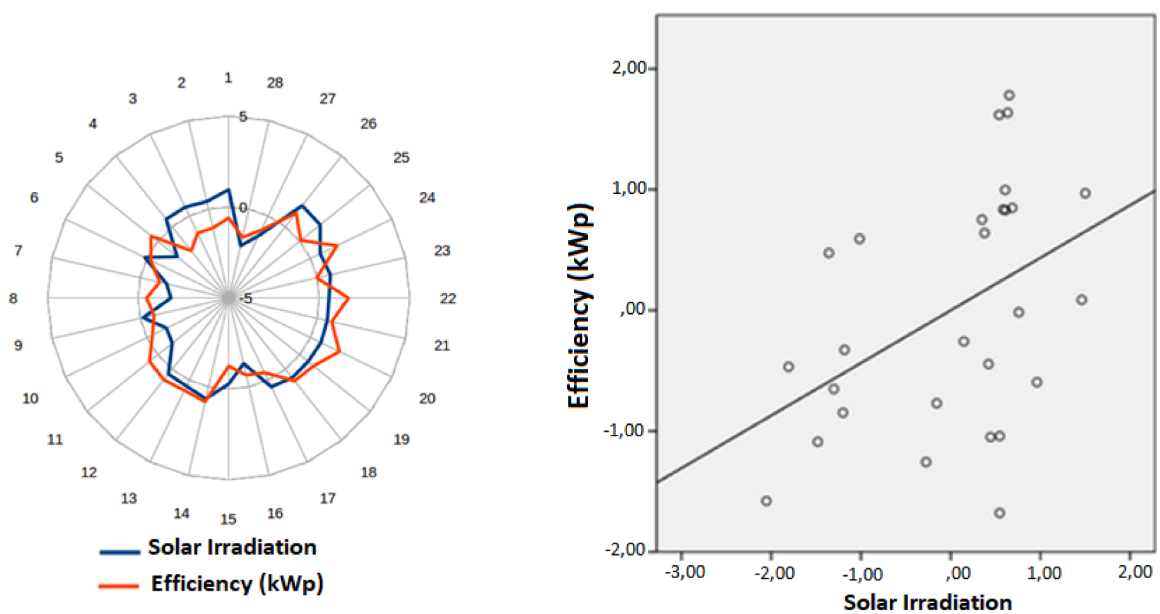
Table 8
Correlations of parameter estimates.

Correlations of parameter estimates								
	b0	b1	b2	b3	b4	b5	b6	b7
b0	1,000							
b1	−,212	1,000						
b2	−,251	−,430	1,000					
b3	,022	,185	−,231	1,000				
b4	−,998	,200	,242	−,052	1,000			
b5	−,095	,222	,123	−,326	,063	1,000		
b6	,097	−,292	−,023	,295	−,103	−,300	1,000	
b7	,131	−,519	,236	,087	−,145	−,111	−,230	1,000



(a) (b)

Fig. 8. (a) Percentage yield and panel efficiency comparison in spider web representation. (b) The scatter plot of Percentage efficiency versus panel yield with Simple Linear Regression. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) (b)

Fig. 9. (a) The comparison of solar irradiation using spider web representation and its relationship with solar panel efficiency. (b) The scatter plot of solar efficiency versus solar irradiation with Simple Linear Regression. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

neural network model outperformed multiple linear regression analysis with an R^2 value of 0.9628. The Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) were found to be 25.09 and 23.89, respectively, for the artificial neural network model. These findings align with previous studies indicating the effectiveness of artificial neural networks in solar panel efficiency analysis. Principal Component Analysis (PCA) was applied to reduce the number of features to two dimensions, and although there was a decrease in performance criteria, it demonstrated that utilizing only two dimensions was sufficient for efficiency prediction. This study highlights the underutilization of artificial neural networks in panel efficiency studies in our country and emphasizes their potential contribution to the energy prediction literature. Furthermore, incorporating weather parameters as input variables could enhance the performance of neural networks for better short-term forecasting. The

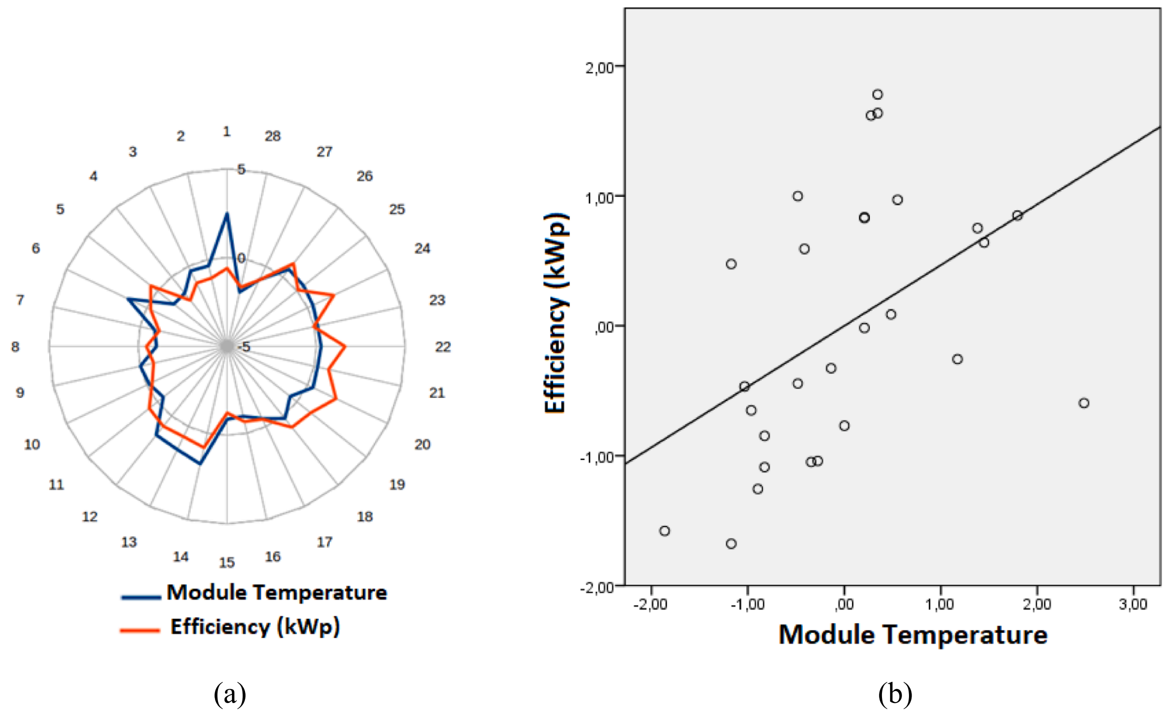


Fig. 10. (a) The solar temperature of spider web representation and panel solar efficiency (b) The scatter plot of solar efficiency versus module temperature with Simple Linear Regression. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

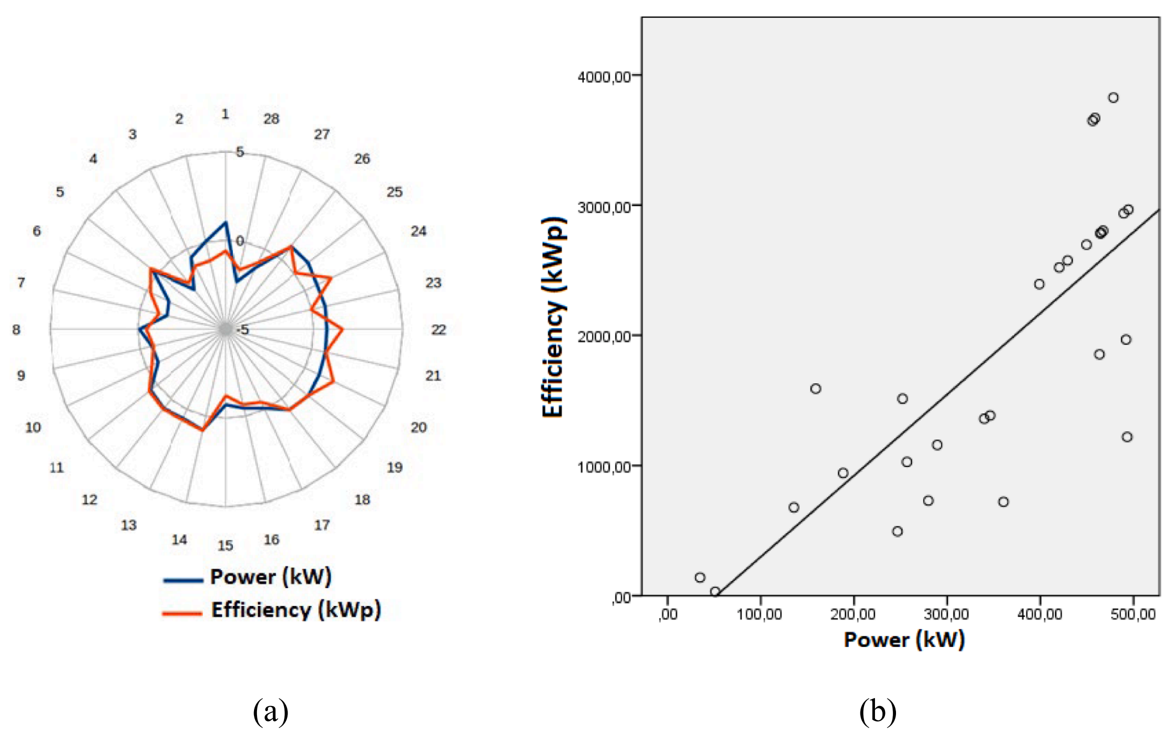
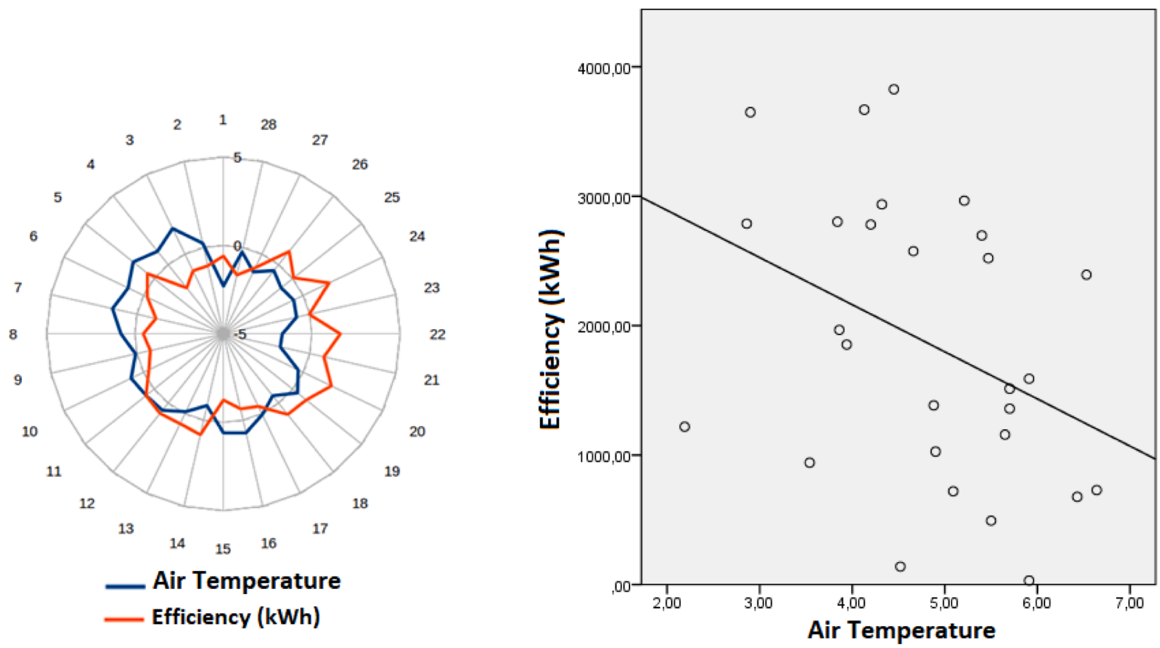


Fig. 11. (a) The solar module power of spider web representation and panel solar efficiency (b) The scatter plot of solar efficiency versus module power with Simple Linear Regression. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

three parameters found to have the greatest impact on the efficiency of the 500 kWp Photovoltaic (PV) Solar Power Plant were solar irradiation, module power, and module temperature. Overall, this study contributes to the understanding and potential optimization of energy investments in the Igdir region and provides insights into the application of artificial neural networks in energy forecasting.

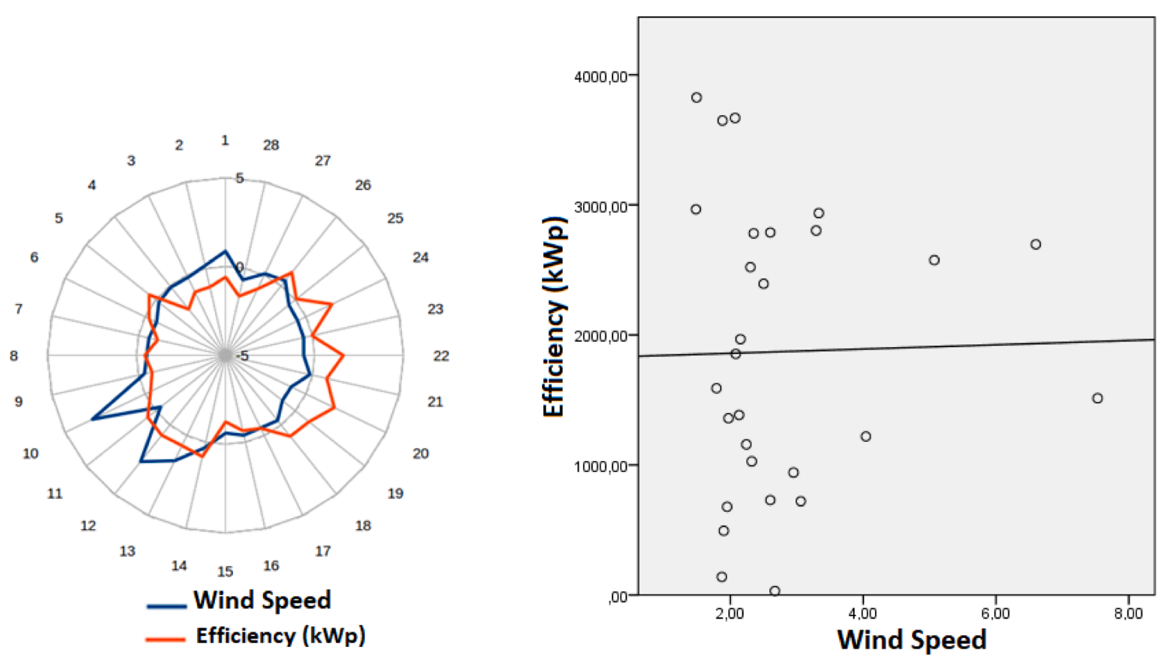
Declaration of competing interest

There is no conflict of interest.



(a) (b)

Fig. 12. (a) The solar efficiency of spider web representation and panel solar efficiency (b) The scatter plot of solar efficiency versus air temperature with Simple Linear Regression. . (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) (b)

Fig. 13. The wind speed of spider web display and panel solar efficiency (b) The scatter plot of solar efficiency versus wind speed with Simple Linear Regression.

Data availability

The authors do not have permission to share data.

References

Abadi, M., Paul Barham, P., Jianmin Chen, J., Zhifeng Chen, Z., Andy Davis, A., Jeffrey Dean, J., Matthieu Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., Brain, G., 2016. TensorFlow: A System for Large-Scale Machine Learning. In: 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). November 2–4. Savannah, GA, USA, ISBN: 978-1-931971-33-1.

Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D., 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew. Sustain. Energy Rev. 124, 109792.

AlShafeey, M., Csáki, C., 2021. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. Energy Rep. 7, 7601–7614.
Bersalli, G., Menanteau, P., El-Methni, J., 2020. Renewable energy policy effectiveness: A panel data analysis across Europe and Latin America. Renew. Sustain. Energy Rev. 133, 110351.
Bhattarai, T.N., Ghimire, S., Mainali, B., Gorjian, S., Treichel, H., Paudel, S.R., 2023. Applications of smart grid technology in Nepal: status, challenges, and opportunities. Environ. Sci. Pollut. Res. 30 (10), 25452–25476.
Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623.
Chollet, F., 2015. Keras, github. https://github.com/fchollet/keras. (Access: May 15, 2023).
Eltamaly, A.M., Mohamed, M.A., 2018. Optimal sizing and designing of hybrid renewable energy systems in smart grid applications. In: Advances in Renewable Energies and Power Technologies. Elsevier, pp. 231–313.
Evangelista, R., Ramalho, E.A., e Silva, J.A., 2020. On the use of hedonic regression models to measure the effect of energy efficiency on residential property transaction prices: Evidence for Portugal and selected data issues. Energy Econ. 86, 104699.
Fentis, A., Bahatti, L., Mestari, M., Chouri, B., 2017. Short-term solar power forecasting using Support Vector Regression and feed-forward NN. In: 2017 15th IEEE International New Circuits and Systems Conference (NEWCAS). Strasbourg, pp. 405–408.
Gensler, A., Henze, J., Sick, B., Raabe, N., 2016. Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM [Neural Networks]. In: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Budapest, 002858-002865.
Ghimire, S., Deo, R.C., Raj, N., Mi, J., 2019. Wavelet-based 3-phase hybrid SVR model trained with satellite-derived predictors, particle swarm optimization and maximum overlap discrete wavelet transform for solar radiation prediction. Renew. Sustain. Energy Rev. 113, 109247.
Heydari, A., Garcia, D.A., Keynia, F., Bisegna, F., Santoli, L., 2019. A novel composite neural network based method for wind and solar power forecasting in microgrids. Appl. Energy 251, 113353.
Hosseini, S.E., Wahid, M.A., 2016. Hydrogen production from renewable and sustainable energy resources: Promising green energy carrier for clean development. Renew. Sustain. Energy Rev. 57, 850–866.
Hwang, H.P.C., Ku, C.C.Y., Chan, J.C.C., 2021. Detection of malfunctioning photovoltaic modules based on machine learning algorithms. IEEE Access 9, 37210–37219.
Ikram, M., Ferasso, M., Sroufe, R., Zhang, Q., 2021. Assessing green technology indicators for cleaner production and sustainable investments in a developing country context. J. Clean. Prod. 322, 129090.
Karatayev, M., Lisiakiewicz, R., Gródek-Szostak, Z., Kotulewicz-Wisińska, K., Nizamova, M., 2021. The promotion of renewable energy technologies in the former soviet bloc: Why, how, and with what prospects? Energy Rep. 7, 6983–6994.
Kaya, F., Sahin, G., Alma, M.H., 2021. Investigation effects of environmental and operating factors on PV panel efficiency using by multivariate linear regression. Int. J. Energy Res. 1–14. https://doi.org/10.1002/er.5717.
Kaytez, F., 2020. A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. Energy 197, 117200.
Khalil, M.I., Jhanjhi, N.Z., Humayun, M., Sivanesan, S., Masud, M., Hossain, M.S., 2021. Hybrid smart grid with sustainable energy efficient resources for smart cities. Sustain. Energy Technol. Assess. 46, 101211.
Khandakar, A., Muhammad, E.H.C., Kazi, M.K., Benhmed, K., Touati, F., Al-Hitmi, M., Antonio, J., Gonzales, S.P., 2019. Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar. Energies 12 (2782).
Kingma, D.P., Ba, J., 2014. A method for stochastic optimization. In: 3rd International Conference for Learning Representations. San Diego (2015).
Koc, A., Turk, S., Sahin, G., 2019. Multi-criteria of wind-solar site selection problem using a GIS-AHP-based approach with an application in Igdir Province/Turkey. Environ. Sci. Pollut. Res. 26 (31), 32298–32310.
Li, L., Wen, S., Tseng, M., Wang, C., 2019. Renewable energy prediction: A novel short-term prediction model of photovoltaic output power. J. Clean. Prod. 228, 359–375.
Liu, Z., Wu, D., Liu, Y., Han, Z., Liyong Lun, L., Gao, J., Jin, G., Cao, G., 2019. Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. Energy Explor. Exploitation 37 (4), 1426–1451.
Maulud, D., Abdulazeez, A.M., 2020. A review on linear regression comprehensive in machine learning. J. Appl. Sci. Technol. Trends 1 (4), 140–147.
Nadimi, R., Tokimatsu, K., 2018. Modeling of quality of life in terms of energy and electricity consumption. Appl. Energy 212, 1282–1294.
Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10). pp. 807–814.
Nazir, M.S., Mahdi, A.J., Bilal, M., Sohail, H.M., Ali, N., Iqbal, H.M., 2019. Environmental impact and pollution-related challenges of renewable wind energy paradigm–a review. Sci. Total Environ. 683, 436–444.
Notton, G., Voyant, C., Fouilloy, A., Duchaud, J., Nivet, M.L., 2019. Some applications of ANN to solar radiation estimation and forecasting for energy applications. Appl. Sci. 9 (1), 209.
Rana, M., Koprinska, I., Agelidis, V.G., 2016. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers. Manage. 121, 380–390.
Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323 (6088), 533–536.
Sahin, F., Işik, G., Şahin, G., Kara, M.K., 2020. Estimation of PM10 levels using feed forward neural networks in Igdir. Turkey. Urban Clim. 34, 100721.
Sharifzadeh, M., Sikinioti-Lock, A., Shah, N., 2019. Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression. Renew. Sustain. Energy Rev. 108, 513–538.
Sharma, N., Sharma, P., Irwin, D., Shenoy, P., 2011. Predicting solar generation from weather forecasts using machine learning. In: 2011 IEEE International Conference on Smart Grid Communications (SmartGridComm). Brussels, pp. 528–533.
Shuvho, M.B.A., Chowdhury, M.A., Ahmed, S., Kashem, M.A., 2019. Prediction of solar irradiation and performance evaluation of grid connected solar 80KWp PV plant in Bangladesh. Energy Rep. 5, 714–722.
Skiba, M., Mrówczyńska, M., Sztubecka, M., Bazan-Krzywoszańska, A., Kazak, J.K., Leśniak, A., Janowiec, F., 2021. Probability estimation of the city's energy efficiency improvement as a result of using the phase change materials in heating networks. Energy 228, 120549.
Sudirman, R., Ashenayi, K., Golbaba, M., 2012. Comparison of methods used for forecasting solar radiation. In: IEEE Green Technologies Conference, Vol. 978. IEEE, TULSA, OK, USA, pp. 673–0967 (1), (2012.04.19-2012.04.20).
Türk, S., Koç, A., Şahin, G., 2021. Multi-criteria of PV solar site selection problem using GIS intuitionistic fuzzy based approach in erzurum province/Turkey. Sci. Rep. 11 (5034).
Varanasi, J., Tripathi, M.M., 2019. K-means clustering based photo voltaic power forecasting using artificial neural network, particle swarm optimization and support vector regression. J. Inf. Optim. Sci. 40 (2), 309–328.
Verma, T., Tiwana, A.P.S., Reddy, C.C., Arora, V., Devanand, P., 2016. Data Analysis to Generate Models Based on Neural Network and Regression for Solar Power Generation Forecasting. In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS). Bangkok, pp. 97–100.
Wang, Z., Koprinska, I., Troncoso, A., Martínez-Álvarez, F., 2018. Static and dynamic ensembles of neural networks for solar power forecasting. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, pp. 1–8.
Wang, J., Zhong, H., Lai, X., Xia, Q., Wang, Y., Kang, C., 2019. Exploring key weather factors from analytical modeling toward improved solar power forecasting. IEEE Trans. Smart Grid 10 (2), 1417–1427.
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30 (1), 79–82.
Yadav, A.K., Chandel, S.S., 2017. Identification of relevant input variables for prediction of 1-minute time-step photovoltaic module power using Artificial Neural Network and Multiple Linear Regression Models. Renew. Sustain. Energy Rev. 77, 955–969.
Zame, K.K., Brehm, C.A., Nitica, A.T., Richard, C.L., Schweitzer III, G.D., 2018. Smart grid and energy storage: Policy recommendations. Renew. Sustain. Energy Rev. 82, 1646–1654.