Energy Conversion and Management 148 (2017) 1070–1081

Contents lists available at ScienceDirect

Energy Conversion and Management

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

Comparison of prediction methods of photovoltaic power system production using a measured dataset

Hussein A. Kazem *, Jabar H. Yousif

Sohar University, PO Box 44, PCI 311 Sohar, Oman

ARTICLE INFO

Article history:
Received 13 April 2017
Received in revised form 20 May 2017
Accepted 18 June 2017
Available online 1 July 2017

Keywords:
Solar energy prediction
Machine learning
Photovoltaic
SOFT-computing
ANN
SVM

ABSTRACT

At present, generating energy from renewable sources is an important topic and is attracting significant attention because of its many benefits. Recent technological developments have made generating renewable energy from various sources such as the sun, wind, geothermal energy, and many other sources a commercially viable process.

In this study, a photovoltaic (PV) system has been designed and installed for energy production. In addition, the PV output was measured for a period of one year. Neural mathematical models such as generalized feedforward networks (GFF), multilayer perceptron (MLP), self-organizing feature maps (SOFM) and support vector machines (SVM) are implemented for simulating and predicting the output of solar energy systems. The practical implementation of the proposed models achieves excellent results in comparison with results found by other researchers. The SOFM attains the lowest MSE value in the training phase (0.0638) compared to the MLP model (0.0645), GFF model (0.0658) and SVM model (0.0693). The accuracy percentage of the proposed models were found to be 80.28% for the GFF and MLP models, 78.55% for the SOFM model and 77.1% for the SVM model. GFF achieved a higher accuracy percentage in comparison with the MPL, SOFM and SVM models and also compared to those found in other studies. The proposed SOFM and MLP models achieved a smaller MAPE value of 5.339 and 5.718 respectively. All of the MLP, GFF, SOFM and SVM models accomplished a low value of RMSE of about 0.25. The proposed models scored excellent NMSE results, especially SVM, which achieved a value of 0.0039.

1. Introduction

Solar energy is considered a very smart and interesting solution to meet the increasing global energy demand because this type of energy is considered to be available on a large scale and in theory will not be exhausted [1]. Solar radiation is the primary source of renewable energy on the Earth. It is very important for solar energy engineers, designers and architects to efficiently identify the needs of water irrigation and crop production, among other things [1,2]. The Sun produces electromagnetic radiation which reaches the Earth and the visible light is only one portion of that radiation. Other invisible radiation includes the Infrared, Ultraviolet UV, radio waves, microwaves, X-rays and gamma rays [3].

Photovoltaic cells (PV), also called solar cells, are devices that create electricity directly from the light that comes from the Sun. PV systems can be utilized in different small, medium and large scale applications [4]. PV technology provides power for many small applications such as lights, TVs and calculators [5]. PV has been available for more than a hundred years [6]. Scientists have tried many different materials to make solar cells to optimize their efficiency. More widely, one of the benefits of PV is that many countries have installed large-scale systems recently to provide consumers with electricity. PV technology began using as a backup systems to provide electrical power to critical equipment and tasks. About 175,000 villages in over 140 countries worldwide use the power that is supported by solar modules, and as a result, thousands of jobs have been created and sustainable economic prospects have been realised [7].

Over 350 megawatts produced by PV systems amounted to over USD 2 billion on the global market in 2001. PV applications have been employed in communications, crop irrigation, healthcare devices, lighting, water purification, environmental monitoring, cathode protection, air navigation and marine systems, utility power, and other commercial and residential applications [7].

Solar radiation and energy data should always be measured precisely over a long period, for the effectual transformation and deployment of solar power. This is due to the importance of solar radiation for PV system design. However, the measurement of solar radiation is sometimes complicated and expensive. Therefore,

* Corresponding author.
E-mail address: h.kazem@soharuni.edu.om (H.A. Kazem).

http://dx.doi.org/10.1016/j.enconman.2017.06.058
0196-8904/

Table 1
Comparison between different modelling techniques.

Authors/reference	Year	Location	Model type	Accuracy	Time consumption	Complexity
Tamer et al. [15]	2012	Malaysia	LM and NLM	LM: Low NLM: Fair	Fast	Low
Modi et al. [16]	1979	India	LM	Low	Fast	Low
Page [17]	1961	40°N–40°S	LM	Low	Fast	Low
Sben et al. [18]	2001	Turkey	LM and NLM	LM: Low NLM: Fair	Fast	Low
Jin et al. [19]	2005	China	LM and NLM	LM: Low NLM: Fair	Fast	Low
Angstrom [20]	1956	USA	LM	Fair	Fast	Low
Altunkaynak et al. [21]	2004	Turkey	MLRM	High	Long	High
Jiang [22]	2008	India	MLRM	High	Long	High
Mehmet et al. [23]	2013	Turkey	ANN	High	Long	High
McCullock et al. [24]	1943	USA	ANN	High	Long	High
Moody et al. [25]	1989	Romania	FL	High	Long	High
Marius et al. [26]	2008	Spain	FL	High	Long	High
Jang [27]	1993	USA	AN-FIM	High	Long	High
Top et al. [28]	1995	Turkey	AN-FIM	High	Long	High
Kennedy [29]	1995	India	PSOM	Fair	Fair	Fair

modelling solar radiation is an important and relatively inexpensive method that can be employed by scientists. For this reason, numerous studies have been proposed in the literature proposing different ways to predict the amount of solar radiation and energy including stochastic prediction models based on time series methods and artificial neural network approaches. Researchers have used different modelling methods like linear and non-linear methods, artificial neural networks (ANN), and fuzzy logic (FL) [6–9].

Modelling methods for solar radiation and energy prediction are in general used to provide information about whether or not solar cells can be deployed in a certain location. For this reason, it is highly important to find the optimum prediction model to match the real data. Many modelling methods are used for this purpose, and in the literature, the most popular models include linear (LM) and non-linear (NLM) modelling, multiple linear regression, hierarchical linear modelling, artificial neural networks (ANN), fuzzy logic (FL), adaptive neuron-fuzzy inference (AN-FIM), and particle swarm optimization (PSOM).

Several researchers are implementing predicting tools for PV system performance [10–14]. These methods in general can be classified as follows: empirical mathematical, regression, and artificial intelligent neural network based models and finally statistical models based on time series of data.

Argiriou et al. [10] implemented a neural controller model based on feed forward back propagation for hydronic heating plants in buildings. The controller was used to forecast variables such as meteorological modules, ambient temperature and solar irradiance. The proposed models rely on real-scale office buildings during real operating conditions. A comparison of operational results and the conventional controller is performed to evaluate the performance of the proposed numerical-based simulation models. The results of the experiments and numerical models showed that the percentage of energy saving is about 15% in North European weather conditions.

Khatib et al. [11] proposed a solar irradiation system-based ANN, which collects data from 28 Cities in Malaysia. The proposed neural network model is used to forecast the clearness index, which is used to establish a predictive global solar irradiation system for Malaysia. The predicted solar irradiation system has a mean absolute percentage error (MAPE) of 5.92%, and yielded 7.96% for the root mean squared error (RMSE) and 1.46% for the MSE.

Dorvloa et al. [12] implemented an ANN technique to estimate the solar radiation-based clearness index. They used long-term data to design a hybrid neural net system-based Radial Basis Function (RBF) and Multi-Layer Perceptron (MLP) for Oman. They claim that both the RBF and MLP models perform well. However, they suggested that RBF models are more efficient since they require less computing power. The RBF is trained with data obtained from different meteorological stations located in Sohar, Seeb, Masirah, Sur, Salalah, Sohar and Fahud.

Sulaiman et al. [13] used hybrid multilayer feed forward neural network (HMLFNN) technique to predict power of grid connected PV system. Artificial immune system (AIS) was selected as an optimized for the multilayer feed forward neural network (MLFNN) training process in the proposed model. The solar irradiance and the module's temperature were used as input variables, while the energy generated by the system was the output of the model. However, a very short period of recorded data were used in this paper for training and testing the proposed HMLFNN technique, which does not consider the annual climate change that consequently affects the prediction accuracy negatively.

Mellit et al. [14] used modelling and simulation of PV power system using adaptive neuro-fuzzy inference scheme (ANFIS) with back propagation (BP) learning algorithm. The ANFIS was developed to model the daily delivered and consumed power generation by the PV power system. Average daily measurements of the metrological data were used as input data for the proposed prediction model, and thus, the uncertainty in solar radiation was not considered.

Table 1 illustrates a comparison between the different modelling techniques in terms of accuracy, time consumption and complexity. The criteria classified are as follows; accuracy (high, low, fair); time consumption (fast, long, fair); and complexity (high, low, fair). A comparison between the mentioned models has been made. ANN has been proven to be accurate and shows better results than multiple linear regression (MLR) as suggested by error index values, making it a very interesting approach. Some authors have claimed that the PSOM model is simple and easy to implement. As Table 1 suggests, ANN modelling techniques are commonly used because of the high degree of accuracy they provide. Therefore, this paper implements an ANN technique to model solar energy production in grid connected PV systems.

2. Solar energy profile of Sohar

Solar radiation and temperature data for a two year period (2011–2012) in Sohar was used. The data contains hourly ambient temperature and global and diffuse solar radiation as shown in Fig. 1. This figure shows that the global and diffused solar energy

for Sohar (daily average) is 6182 W h/m^2 and 3289 W h/m^2, respectively. In other words, the Sohar area can potentially be invested in with regard to PV and thermal solar systems. However, the high ambient temperature, dust and humidity need to be taken into consideration. Sohar's daily average ambient temperature is 32 °C and the effect of dust on PV performance is less compared with other countries in the Gulf region [19]. Finally, the monthly average humidity is 72.5% [20].

3. Grid connected PV system model

Grid connected PV systems contain PV arrays and an inverter. A10 PV modules have been installed in Sohar University in Oman. The rating of the PV module is 140 W, 17.7 V maximum voltage, 7.91 A maximum current, 22.1 V open circuit voltage, 8.68 A short circuit current, and 13.9% efficiency as shown in Fig. 2. In addition, an inverter with the following specifications have been installed: 1.7 kW rated power, 220–240 V AC voltage, 50 Hz and 94.1% efficiency.

3.1. Modelling of PV array

The available "solar radiation" (G) and "ambient temperature" (T) are the keys to determining the output power of a PV array. The PV array output power increases and decreases linearly as the solar radiation and ambient temperature increases, respectively. Thus, the instantaneous output power of a PV array [15] can be given by,

$$P_{PV}(t) = P_{Peak}\left(\frac{G(t)}{G_{standard}}\right) - \alpha_T[T_c(t) - T_{standard}] \quad (1)$$

where $G_{standard}$ and $T_{standard}$ are the standard test conditions STc for solar radiation and cell temperature, respectively and α_T and P_{peak} are the temperature coefficient and rated power of the PV, respectively. The cell temperature (Tc) is defined as in equation [6],

$$T_C(t) - T_{ambient} = \frac{T_{standard}}{800}G(t) \quad (2)$$

Maximum power point trackers (MPPT) are used to extract the maximum power under varying weather and load conditions, which is the optimum operation of the PV. Based on Eqs. (1) and (2), the PV array power has been calculated.

3.2. Inverter model

The efficiency of an inverter [8] can be calculated as in Eq. (3).

$$\eta(t) = \frac{P_{in}(t) - P_{Loss}(t)}{P_{in}(t)} \quad (3)$$

where $P_{in}(t)$ and $P_{Loss}(t)$ are the instantaneous input and loss power, respectively.Because of the losses in the system is not constant but depends on many conditions which make it difficult to be calculated. Therefore, it is important to develop an inverter model for the inverter efficiency.

Fig. 3 shows an efficiency curve for a commercial inverter obtained from the datasheet. The curve describes the inverter's efficiency (in percent) regarding input power and inverter rated power.

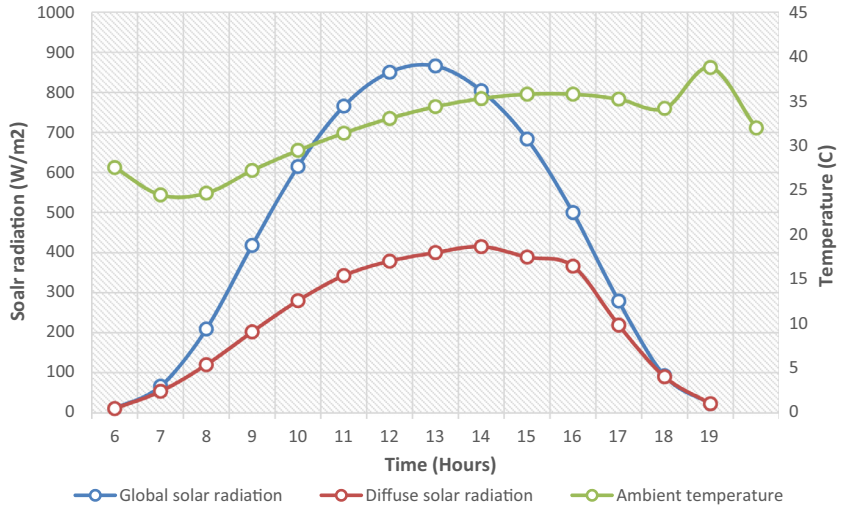

Fig. 1. Sohar solar radiation and ambient temperature profile.

Fig. 2. Capture of the installed PV system.

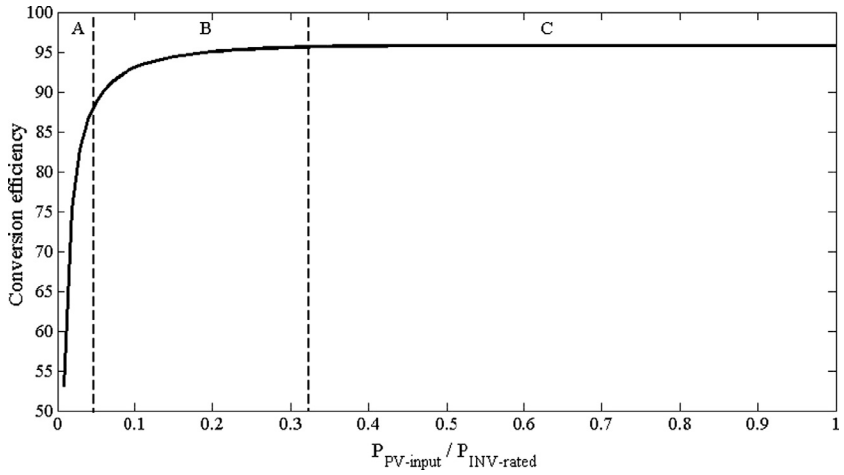

Fig. 3. Inverter typical efficiency curve.

A power function can describe the efficiency curve as follows [30],

$$\begin{cases} \eta = c_1\left(\frac{P_{PV_{input}}}{P_{INV_{Rated}}}\right)^{c_2} + c_3 & \frac{P_{PV_{input}}}{P_{INV_{Rated}}} > 0 \\ \eta = 0 & \frac{P_{PV_{input}}}{P_{INV_{Rated}}} = 0 \end{cases} \quad (4)$$

where P_{PV} and P_{INV_C} are output power of the PV system and rated power of the inverter, respectively, while C1–C3 are the model factors.

4. Grid connected PV system evaluation criteria

To evaluate the grid-connected PV system, one technical and two economic criteria have been used. The technical criteria are the capacity and yield factors. The former evaluates the usage of the PV system and the latter assists the productivity of the PV system. On the other hand, the economic criterion is the cost of energy produced (*CoE*). The yield factor (*YF*) is defined as "the annual, monthly or daily net AC energy output of the system divided by the peak power of the installed PV array at standard test conditions (STC)". It is given by [15,30],

$$YF = \frac{E_{PV}\ (\text{kW h/year})}{PV_{WP}\ (\text{kWp})} \quad (5)$$

The capacity factor (CF) is defined as the "ratio of the actual annual energy output to the amount of energy the PV array would generate if it operated at full rated power (Pr) for 24 h per day for a year". It is given by [8],

$$CF = \frac{Y_F}{8760} = \frac{E_{PVannual}}{P_R * 8760} \quad (6)$$

This economic criterion simulates the unit cost, life-cycle cost and payback period. The life cycle cost is given by [6],

$$LCC = C_{capital} + \sum_1^n C_{O\&M} + \sum_1^n C_{replacment} - C_{salvage} \quad (7)$$

where $C_{capital}$ is the capital, $C_{replacement}$ is the replacement, and $C_{O\&M}$ is the operation and maintenance costs of the system, respectively. Finally, the salvage value ($C_{salvage}$) of a system is its net worth in the final year of the life-cycle period. After calculating the LCC, the unit cost of the energy produced by the system can be calculated as in Eq. (8).

$$CoE = \frac{LCC}{\sum_1^n E_{PV}} \quad (8)$$

5. Neutral networks modelling

Neural computing (NC) implements computational methodologies for imitating the activities of biological and physical systems. It includes a range of techniques and methods which aim to discover a mathematical model for such complex and dynamic systems for real-life problems. These problems include data classification and filtering, data association and mining, pattern and image recognition, NLP, etc. [31]. Neural networks are biologically inspired mathematical models that have the ability to learn and disseminate data for both linear and non-linear systems. Neural networks have certain advantages such as their capability to evaluate, decide, check, and calculate within a precise and imprecise domain. Furthermore, they can emulate the human ability to learn from experience. The computation of neural networks includes several phases such as training, learning, testing and generating the PV output, and then generalized the results for unseen data [32]. Generalized feedforward networks (GFF) are a generalization of the MLP such that connections can jump over one or more layers. However, GFF often solves the problem much more efficiently. Standard MLP requires hundreds of times more training epochs than the generalized feedforward network containing the same number of processing elements. MLP is one of the most used techniques, and is used in neural computations because it can easily be applied in many applications. The MLP neural network is a feed-forward supervised approach because it requires input and output data to achieve the desired values. SOFM is an unsupervised neural computation method based on the concept of feature maps. Feature maps can systematize the input space of high-dimensional data into two-dimensional nodes. The SOFM learning process changes the weight of winner nodes and its neighbourhood closer in the input data space. Usually, the adjustment of the weighted values is implemented using the Hebbian rule [33].

Support Vector machines (SVM) are training algorithms for the learning classification and regression rules from the data [34]. However, SVM systems implement linear algorithms in a high dimensional feature space to train the input data sets. The learning algorithm applies a polynomial and radial basis function (RBF). SVMs in NeuroSolutions are a new kind of classifier incited by two connotations. First, is to maps data into a high-dimensional space that converts non-linear data into linear data. Second, the training SVM algorithm uses only inputs close to the decision surface since they provide the most information about the classification.

6. Results and discussions

6.1. Neural computing

Neurosolutions software was used for designing and implementing the neural networks to simulate and predict the renewable energy power of the PV system for the area of Sohar in the Sultanate of Oman. The same conditions apply to the proposed models to compare the results efficiently. This paper implements the same data sets (5000) to create and train the neural network models. These types apply the same number of input, hidden and output layers and transfer functions. The maximum number of epochs is 1000. This paper includes the implementation of several types of neural cells, including GFF, MLP, SOFM and SVM. The input variables are solar radiation and ambient temperature, while the output is PV voltage. The TanhAxon transfer function implemented in hidden and the output layer. The TanhAxon function applies a bias and Tanh formula to each neuron in the layer, which flattens the range of each neuron between −1 and 1. Such non-linear elements provide a network with the ability to make soft decisions. The back propagation-learning algorithm (BP) propagates the errors and adapts the weights in the hidden PEs. The BP learning method with the error function is defined as in Eq. (9).

$$E(w) = \sum_{p=1}^{pt} \sum_{i=1}^{Epoch} (di(p) - yi(p))^2 \tag{9}$$

where $E(w)$ is the error function to be minimized, w is the weight vector, pt is the number of training patterns, $Epoch$ is the number of output neurons, and $di(p)$ is the desired output of neuron i. $yi(p)$ is the actual output of the neuron i. Table 2 shows the summary of these conditions.

The architecture of the neural network organizes the neurons into layers with a suitable connection configuration between them. Furthermore, the weight of each neuron is initialized with random numbers. The training of ANN involves the adjustment process of weights to fit the training data set. The ANN can then generalize to learn unseen samples. The encoding of an input text helps to transfer the entry data into a suitable digital form. The ANN will categorise input data into three types, the training, cross-validation and testing data sets. The cross-validation data sets adjust the weights of the network and the test data sets for testing the output of the proposed model. The comparison of cross-validation data sets with training datasets helps to determine the error in the results of the network (see Fig. 4).

The momentum rule applies for learning algorithm with a step size equal to 1 and a rate equal to 0.7. The genetic optimization helps to find the best values for the parameters of the momentum rule. There are several techniques employed to determine the performance of the network such as the mean squared error (MSE), which is equal to two times the average cost. The MSE is defined as in Eq. (10).

$$MSE = \frac{\sum_{j=o}^{P} \sum_{i=0}^{N} (d_{ij} - y_{ij})^2}{NP} \tag{10}$$

The summation of squared error (SSE) is specified by the number of population N. SEE is represented as in Eq. (11).

$$SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2 \tag{11}$$

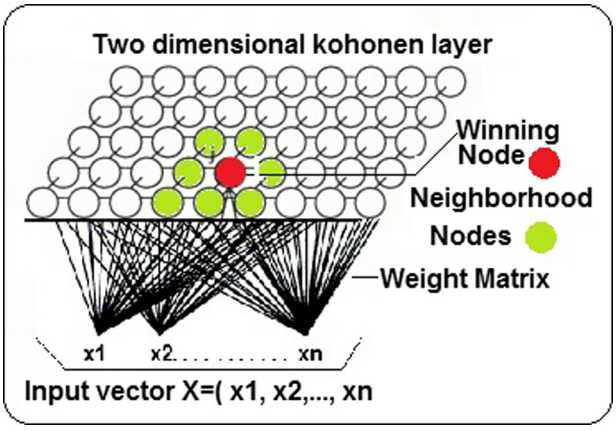

Fig. 4. SOFM implementation panel using NeuroSolution.

The value of output processing elements is P and N is the number of examples in the training data set. y_{ij} is the network output, for example, i at processing element j. d_{ij} is the desired output for example i at processing element j.

NMSE (normalized mean square error) is an estimator of the overall deviations between the predicted and measured values which determines the striking differences among the models. Therefore, the model is performing well if it has a very low NMSE value. In contrast, a high NMSE value means that a model performs less well. NMSE is determined as in Eq. (12).

$$\text{NMSE} = \frac{\text{P} \quad \text{N} \quad \text{MSE}}{\sum_{\text{j=o}}^{\text{P}} \frac{\text{N}\sum_{\text{i=0}}^{\text{N}} \text{d}_{\text{ij}}^2 - \left(\sum_{\text{i=0}}^{\text{N}} \text{d}_{\text{ij}}\right)^2}{\text{N}}} \tag{12}$$

Mean absolute error (MAE) is the average deviation of all absolute errors. Usually, a lower MAE value means a better prediction. MAE is determined as in Eq. (13).

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |d_{ij} - y_{ij}| \tag{13}$$

Root-mean-square error (RMSE) determines the differences between the values of the predicted model and the actual values of real models. RMSE is defined as in Eq. (14).

$$RMSE = \sqrt{\frac{1}{n} \sum\nolimits_{i=1}^{n} (d_i - y_i)} \tag{14}$$

The mean absolute percentage error (MAPE) determines the percentage of error between the predicted value and the desired value. MAPE can only be useful for non-negative values. MAPE is described as in Eq. (15).

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{y_i - d_i}{y_i} \right| \tag{15}$$

Accuracy (AC) is the degree of closeness of results in comparison to the actual values. AC is defined as in Eq. (16).

$$\text{AC} = \frac{(A + D)}{(A + D + B + C)} * 100 \tag{16}$$

A is the negative number of the correct predictions.
B is the positive number of the incorrect predictions.

Table 2
summarized the conditions of ANN.

Factors	Number of input layers	Number of hidden layers	Number of output layers	Transfer function	Learning Rule	Learning Rate	Step size
No. type	2	1	1	TanhAxon	Momentum	0.7	0.1

C is the negative number of the incorrect predictions.
D is the positive number of the correct predictions.

The discussion of the results will address different performance factors which measure the effectiveness of each model. Some of these factors include MSE and RMSE and the accuracy will test how the proposed models will fit the desired output. On the other hand, other factors such as MAE, MAPE and R^2 verify how the prediction values will fit the desired output. Fig. 5 illustrates the MSE for the proposed neural models.

Fig. 5 shows that SOFM achieves a small MSE value in the training phase equal to 0.0638. Moreover, the MLP model achieves an MSE value equal to 0.0645. The GFF model achieves an MSE value equal to 0.0658 and lastly, the SVM model yields an MSE value equal to 0.0693. Moreover, Fig. 6 illustrates the accuracy percentage of the proposed models compared with Ref. [22], which has a high proportion of 98%. On the other hand, the GFF model achieves 80.28% and MLP and SOFM achieve 78.55%. Moreover, SVM achieves an accuracy of 77.1%. The proposed models yielded relatively similar accuracy ratios. GFF produces a high accuracy percentage in comparison with MPL, SOFM, and SVM (see Fig. 7).

MAPE helps to verify if the predicting model results fit the results of the practical design or not. Usually, smaller values for MAPE indicate a better fitting model. Therefore, the MLP prototype in Refs. [39,30] achieve lower values of 3.29 and 4.49, respectively. However, they are implementing different data sets for different environments. The proposed models apply the same conditions as those in Table 2. The SOFM and MLP achieve a smaller value of 5.339 and 5.718 respectively.

Fig. 8 depicts a comparison of the proposed models and other studies based on RMSE. All of the proposed MLP, GFF, SOFM and SVM models achieve a smaller value of about 0.25. On the other hand, other MLP models such as those in Refs. [10,36,38] and the SVM model in Ref. [39] obtain high RMSE values of between 5 and 10. Figs. 9–12 illustrate that the desired output and actual output of the proposed MLP, GFF, SOFM and SVM models are well-fitting and predicting.

Table 3 shows the performance comparison of the proposed models based on different factors such as MSE, RMSE, MAPE, NMSE, MAE, minimum absolute error, accuracy and correlation. Most of the proposed models achieve excellent results. The correlation factor proved that there was a strong relation between input and output variables, which indicates excellent fitting of the proposed model outputs and the desired outputs. NMSE shows that the overall deviations between predicted and measured values are in close result around 0.2, which control the striking differences among proposed models. Moreover, the minimum absolute error is used to compute the differences between the desired output and the predicted results. The proposed models scored excellent results, especially SVM, which achieves a value of 0.0039.

The one-sample *t*-test is implemented to provide more mathematic details for the simulated results in Figs. 9–12. The one-sample *t*-test decides if the mean of samples follows a normal distribution or not for some possible value for a given confidence

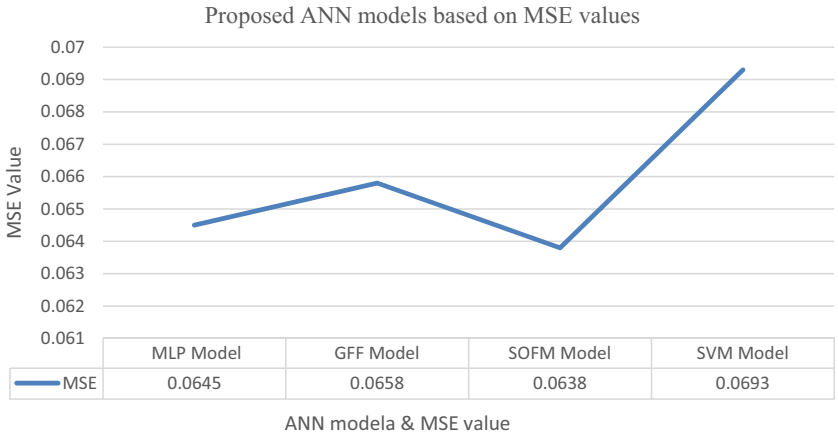

Fig. 5. Proposed ANN models based on MSE values.

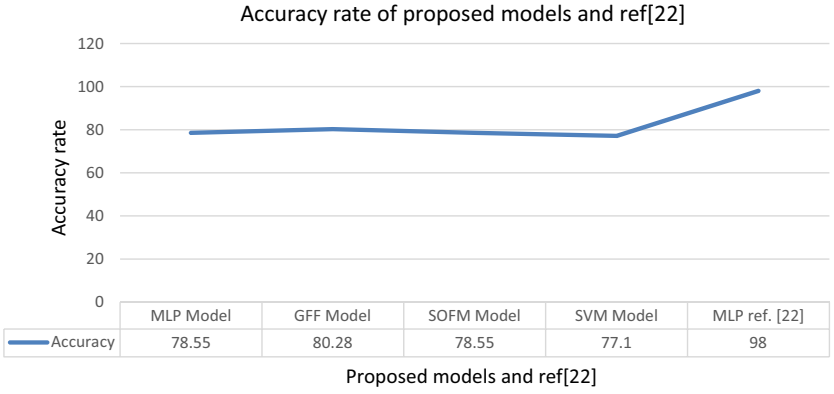

Fig. 6. Proposed ANN models based on accuracy values.

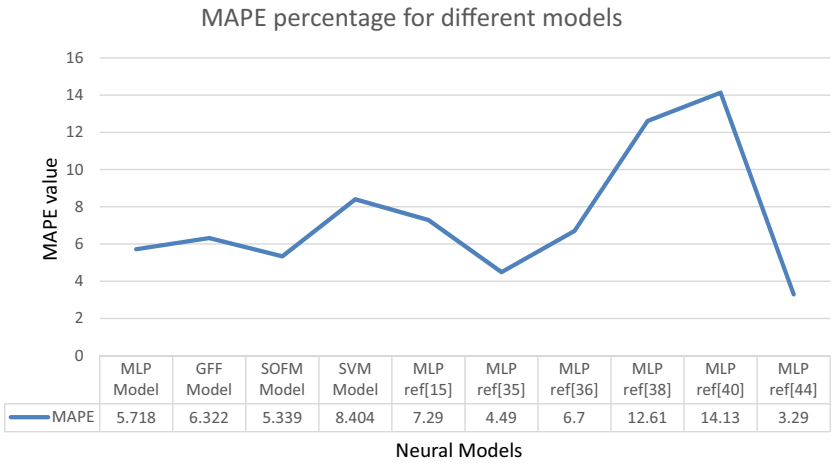

Fig. 7. The MAPE percentage for different models.

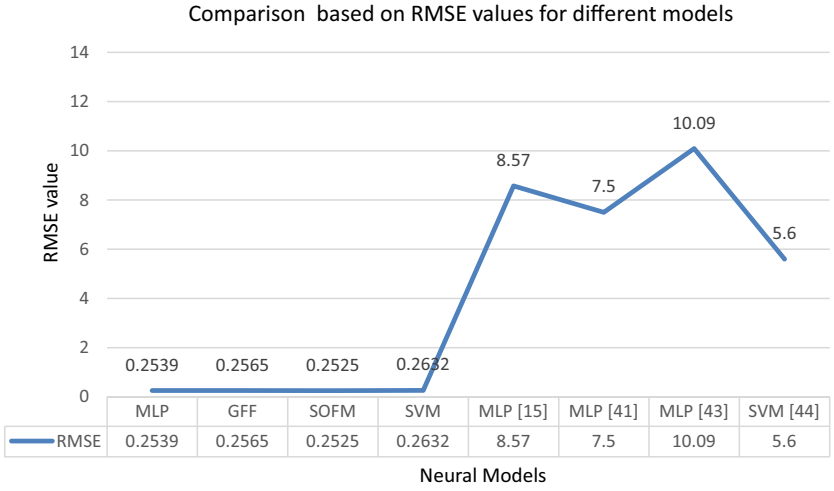

Fig. 8. Comparison based on RMSE values for different models.

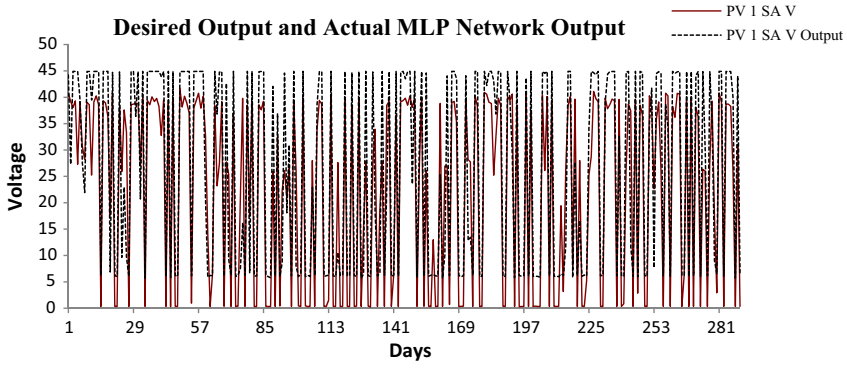

Fig. 9. The desired output and actual MLP network output.

level. The values of *t* statistic and p-value determines either accept or reject the null hypothesis with a small p-value. The p-value defines the probability value of null hypothesis is true. Table 4 concludes the result of one sample test using Origin Lab software. The factors of the table are sample size (N) is 290 samples (days) and standard error mean (SEM), the value of t statistics, degree of freedom (DF) and the value of Prob > |t|.

The mean value of the desired model is 22.45. Therefore, one sample *t*-test will examine if the proposed models follow the normal distribution and to determine if accepts or reject the null hypothesis.

The Null Hypothesis said Mean = 22.45, while the Alternative Hypothesis said Mean <> 22.45.

The result shows that at the 0.05 significance level, all the proposed models are significantly different from the test mean (22.45)

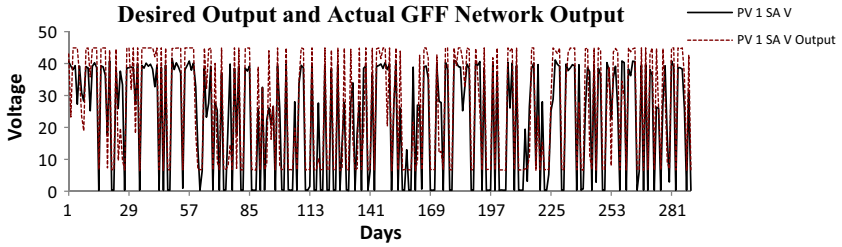

Fig. 10. The desired output and actual GFF network output.

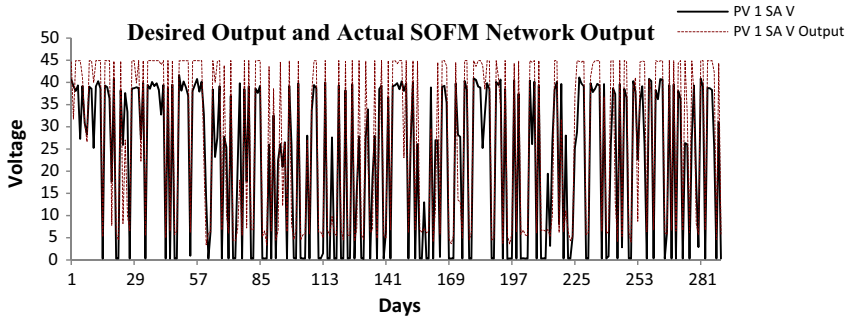

Fig. 11. The desired output and actual SOFM network output.

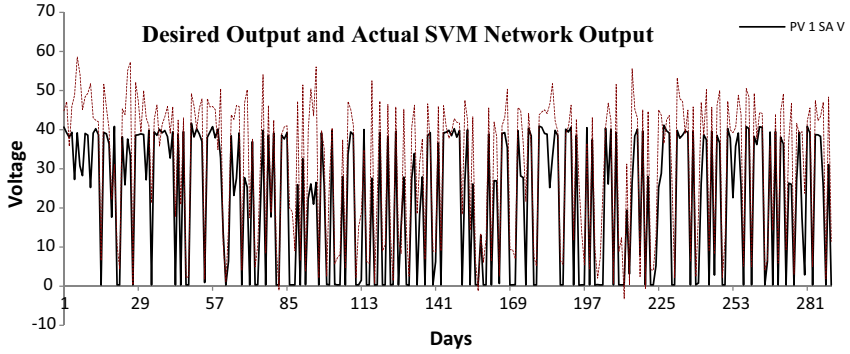

Fig. 12. The desired output and actual SVM network output.

Table 3
the performance comparison of the proposed models based on different factors.

Performance	MLP model	GFF model	SOFM model	SVM model
MSE	0.0645	0.0658	0.0638	0.0693
RMSE	0.2539	0.2565	0.2525	0.2632
MAPE	5.718	6.322	5.339	8.404
NMSE	0.2355	0.2497	0.2277	0.4114
MAE	7.2593	7.5038	7.0096	8.7785
Min Abs Error	0.0305	0.1413	0.0473	0.0039
Accuracy	78.55	80.28	78.55	77.1
r	0.9109	0.8965	0.9181	0.9118

Table 4
the one-sample *t*-test results for desired and proposed models.

Model	N	SEM	t Statistic	DF (N-1)	Prob > |t|
Desired	290	1.021	0.006	289	0.99503
MLP	290	1.067	3.525	289	0.0004
SVM	290	1.061	7.794	289	0
SOFM	290	1.085	3.566	289	0.0004
GFF	290	1.039	3.286	289	0.0011

because the value of Prob > |t| is less than 0.05. Then the null hypothesis is rejected and accepts the alternative hypothesis. The result of hypothesis as follows.

Desired model: At the 0.05 level, the population mean is NOT significantly different from the test mean (22.45).
MLP model: At the 0.05 level, the population mean is significantly different from the test mean (22.45).
SVM model: At the 0.05 level, the population mean is significantly different from the test mean (22.45).
SOFM model: At the 0.05 level, the population mean is significantly different from the test mean (22.45).
GFF model: At the 0.05 level, the population mean is significantly different from the test mean (22.45).

Even with the current-status of PV, many challenges continue to face this technology which can be summarized as degradation due to environmental elements. A reduction of 5% in efficiency is resultant of an increase of cell temperature of 10 °C [45]. The sporadic productivity and uncertainty of PV is a clear issue. Changing of solar radiation causes fluctuations in the output, which might deem the technology unreliable [46–47]. Thus, the instantaneous output power of a PV array [30] can be given by,

$$P_{PV}(t) = P_{Peak}\left(\frac{G(t)}{G_{standard}}\right) - \alpha_T[T_c(t) - T_{standard}] \quad (17)$$

where $G(t)$, $G_{standard}$, T_c and $T_{standard}$ are the solar radiation, standard test conditions for solar radiation, ambient temperature and cell temperature, respectively and α_T is the temperature coefficient of the PV module power which can be obtained from the datasheet. Eq. (17) shows that solar radiation and temperature are the main variables, which specify the PV power production.

Ref. [48] summarized various correlations between temperature and PV efficiency proposed in the literature. However, the authors claimed that careful in applying a particular expression for the operating temperature of a PV module is important because the available equations have been developed with a specific mounting geometry or building integration level in mind. The sensitivity analysis evaluates the relative importance of explanatory variables, with a few significant distinctions. The sensitivity analysis determines the relationship between variables rather than a categorical description. A high value of sensitivity analysis means a strong association between the variables. Fig. 13 depicts the sensitivity analysis test of ambient temperature and solar radiation for different networks based on the experimental data of the system in Fig. 2. From the results in Fig. 13, it is clear that solar radiation is the most important factor influencing the output in the GFF, MLP, SOFM and SVM networks comparing with temperature. This result is important since it shows that the impact of solar radiation in more on the PV power productivity than the temperature. This indicates that PV power production is directly proportional to solar radiation and inverse proportional to temperature. Since both variables are high in Oman, the sensitivity analysis indicates that the relatively high temperature is not a barrier to PV performance compared with the relatively high solar radiation.

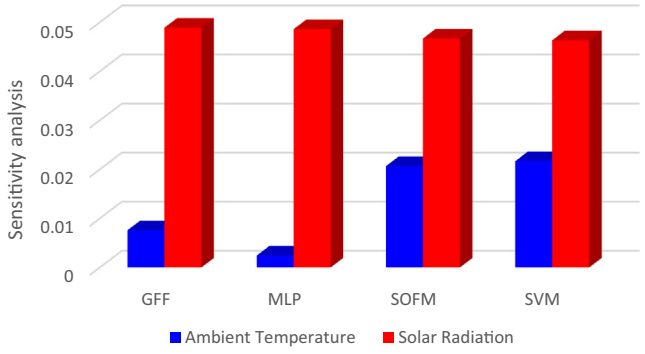

Fig. 13. Sensitivity analysis test for ambient temperature and solar radiation.

6.2. Prediction of mathematical models

The coefficient of determination (R^2) is used to evaluate the validity of the results obtained from predictive results and to match them to the actual results derived from the real model. However, in some cases, R^2 yields unclear results, especially with time-series analysis. Therefore, MAPE and MAE are applied to verify the validity of the forecasting results. For this purpose, we will use the regression models to create a new installation curve using the 2-D Curve Table software. In addition, the "residual" calculation method is used to identify the difference between the actual results and the prediction results. Hence, the value of the residues is assumed to be in the range of a normal distribution with an average of zero and a specified standard deviation (SD) in the ordinary least squares regression. The residual standard error (RSE) approximates the standard deviation and significantly represents the variability in the dependent variable of the proposed model as calculated by Eq. (17). A lower value of RSE indicates a better prediction model.

$$RSE = \sqrt{1 - R^2}SD \quad (18)$$

Since $R^2 > 0$, then RSE will have a value lower than the SD. Moreover, in rare cases, when $R^2 = 0$, then RSE will equal the value of the SD. Fig. 14 shows that the best fits must be in the range of [0,5.04]. The SOFM model obtains results in the range of the best fit and the fitting curve is shown in Fig. 15.

Fig. 15 shows the fitting curve of the SOFM model using linear regression in which an R^2 equal to 0.6848 and SD equal to 4.90 are obtained. In addition, the MLP model obtained R^2 equal to 0.6258 and an SD equal to 5.03. Moreover, Fig. 16 illustrates the residual curve for the MLP output, which indicates that most of the output is in the range of (0,5.03).

6.3. Comparison with similar works

Table 5 shows a comparison of the experimental results of the four models (MLP, GFF, SOFM, SVM) for the standalone PV system voltage (SA V). The mean value, which refers to one measure of the central tendency, shows that GFF more closely resembles the experimental results, while SVM does less so but within the acceptable range. The standard deviation shows that GFF more closely resembles the experimental results, which quantifies the amount of variation model values. On the other hand, of the four models, SEM seems to more closely resemble the experimental results, which depicts the relationship between the dispersion of individual observations around the standard deviation.

To measure the "tailedness" and "asymmetry" of the probability distribution of a real-valued random variable Kurtosis and Skewness are used, respectively. However, it was found that for Kurtosis, GFF is closer, while SVM is closer for Skewness. There is no consistency found in the maximum values of the four models. However, for minimum MLP, GFF, and SOFM are consistent with the experimental results.

To validate the model results, a comparison is shown in Table 6 of different ANN techniques applied in different studies in the literature for the modelling or prediction of PV energy and solar radiation in the period 1998–2016. The general error of these models was found to be in the range of 3% and 14% for monthly, daily,

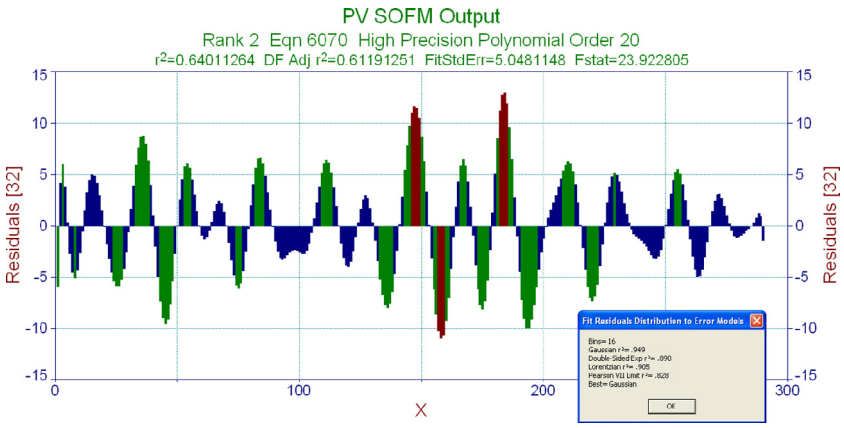

Fig. 14. The residual curve for the SOFM output.

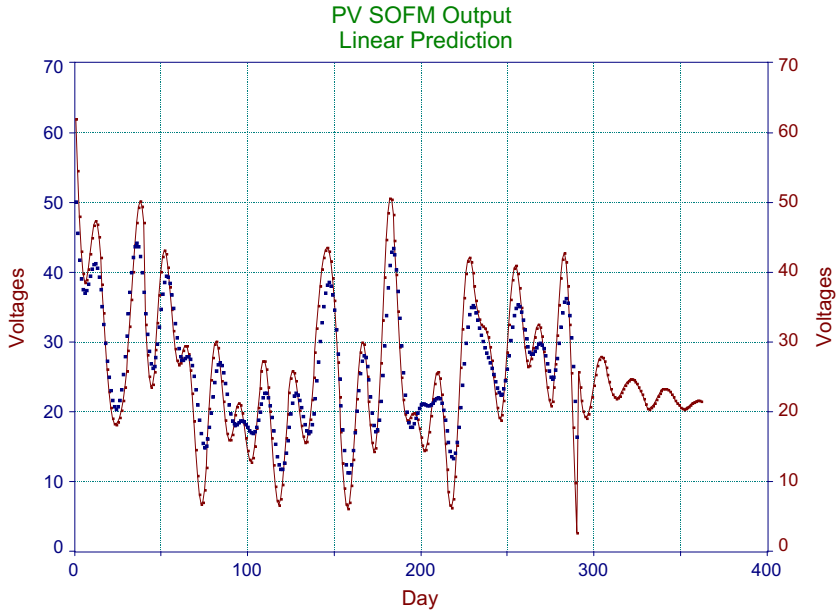

Fig. 15. The fitting cure of SOFM model using linear regression.

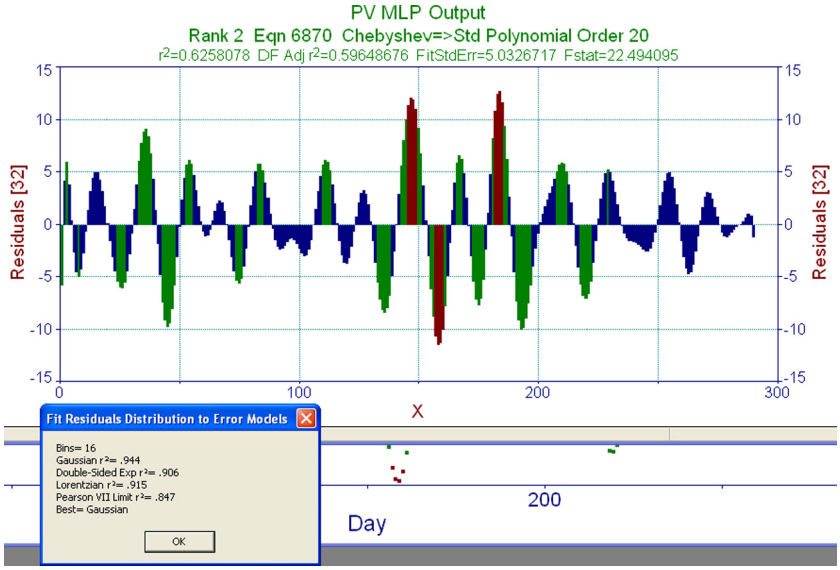

Fig. 16. The residual curve for the MLP output.

Table 5
Comparison of the statistic for the experiment results and the four models (MLP, GFF, SOFM, SVM).

Factor/variables	Experimental	MLP	GFF	SOFM	SVM
Mean	22.4564	26.214	25.868	26.322	30.720
Standard Deviation	17.388	18.184	17.709	18.486	18.069
Standard Error of Mean (SEM)	1.021	1.068	1.039	1.086	1.061
Median	27.669	30.901	26.579	33.356	39.672
Kurtosis	−1.716	−1.908	−1.898	−1.911	−1.4159
Skewness	−0.330	−0.085	−0.015	−0.121	−0.488
Range	41.393	39.245	38.484	41.554	61.844
Minimum	0.241	5.671	6.432	3.358	−3.239
Maximum	41.634	44.915	44.916	44.912	58.605
Sum	6512.347	7602.077	7501.603	7633.307	8908.871

Table 6
Comparison of production accuracy results.

Authors/reference	Year	Location	Network Type	Model
Tamer et al. [15]	2012	Malaysia	MLP	MBE: −1.12% RMSE: 8.57% MAPE: 7.29%
Jiang et al. [22]	2009	China	MLP	Accuracy 95%
Rehman et al. [35]	1998	Saudi Arabia	MLP	MAPE: 4.49%
Sozen et al. [36]	2005	Turkey	MLP	MAPE: 6.7% R value: 0.958
Elminir et al. [37]	2009	Egypt	MLP	SD: 9.00%
Mohandes et al. [38]	2008	Saudi Arabia	MLP	MAPE: 12.61%
Mubiru et al. [39]	2008	Augenda	MLP	MAPE: 7.3% R value: 0.960
Bilgili et al. [40]	2007	Turkey	MLP	MAPE: 14.13% R value: 0.960
Bosch et al. [41]	2008	Spain	MLP	RMSE: 7.5%
Fadare [42]	2009	Nigeria	MLP	SD: 11.04% R value: 0.978
Graditi et al. [43]	2016	Italy	MLP	RMSE: 10.09% R2: 0.978
Chao et al. [44]	2016	USA	SVM	MAPE: 3.29% RMSE: 5.60% R value: 0.998
Current study	2017	Oman	MLP GFF SOFM SVM	MSE: 0.0645, 0.0658, 0.0638, 0.0693 RMSE 0.2539, 0.2565, 0.2525, 0.2632 MAPE 5.718, 6.322, 5.339, 8.404 MAE: 7.2593, 7.5038, 7.0096, 8.7785 R value: 0.9109, 0.8965, 0.9181, 0.9118 Accuracy: 78.55%, 80.28%, 78.55%

hourly and minute predictions. Moreover, MLP, SVM and SOFM are the main techniques employed to model or predict PV energy and solar radiation. However, it is noted that MLP is the most commonly used technique and can be used with exogenous parameters or coupled with other predictors.

Table 3 illustrates the accuracy of production models based on factors such as MAPE, MBE, RMSE, MAE, and R to verify the results of the proposed models. In terms of R and RMSE or MSE, the ANN model best fits the experimental data. The best results of the proposed neural model achieved an MSE value of 0.0638. The R value of the proposed models is similar to the results in [36,40,42,44]. However, the accuracy of the GFF model (80.28%) is the highest among the models investigated.

7. Conclusions

The aim of this study was to propose ANN and mathematical models for the electrical energy production of a photovoltaic system using ANN. However, energy prediction using ANN is not new and many researchers have used and implemented neural models previously [10–39].

It is worth noting that most of these applications are based on the use of supervised feed-forward ANN. On the other hand, few research papers have focused on using recurrent ANN and unsupervised models. Therefore, the first contribution of this paper is to propose four neural computing techniques namely GFF, MLP, SOFM and SVM are based on supervised neural network. The proposed SOFM model is based on unsupervised feedback neural network for improving the performance of the network. Solar radiation and ambient temperature data were the inputs of these models, while the PV array voltage and the current were the outputs. A full year of experimental data for a system located in Sohar, Oman was used. The four models were compared in terms of MSE, NMSE, MAE, accuracy and R. SOFM yielded the smallest MSE value in the training phase equal to 0.0638, while the MLP, GFF and SVM models yielded values of 0.0645, 0.0658 and 0.0693, respectively. Moreover, the accuracy percentage of the models was 80.28% (GFF), 78.55% (MLP, SOFM) and 77.1% (SVM). As can be seen, GFF yielded a higher accuracy percentage in comparison with MPL,

SOFM, and SVM. The proposed SOFM and MLP models gave smaller MAPE values of 5.339 and 5.718 respectively. All four models yielded a small value of RMSE of about 0.25. The proposed models scored excellent NMSE results, especially in the case of SVM which achieved a value of 0.0039.

Moreover, sensitivity analysis was implemented and it was found that PV production is more sensitive to solar radiation than ambient temperature. Finally, the proposed models were compared in terms of statistics. It was found that the GFF model fits the experimental results best. However, for minimum MLP, GFF, and SOFM are consistent with the experimental results. Moreover, the comparison of the results of the proposed models with related work for the purpose of validation of the production results show that the proposed models are encouraging, as depicted in Tables 3, 5, and 6.

Conflict of interests

"The authors declare that there is no conflict of interests regarding the publication of this paper".

Acknowledgment

The research leading to these results has received Research Project Grant Funding from the Research Council of the Sultanate of Oman, Research Grant Agreement No. ORG SU EI 11 010. The authors would like to acknowledge support from the Research Council of Oman.

References

[1] Hill R. Prospects for photovoltaics. Energy World 1999; 208 [original data updated by Hynes K. and Hill R, 8–11].
[2] Aslania Alireza, Wong Kau-Fui V. Analysis of renewable energy development to power generation in the United States. Renew Energy 2014;63:153–61.
[3] CSIRO, BOM. State of the climate. Australia: Commonwealth Scientific and Industrial Research Organisation (CSIRO) and the Bureau of Meteorology (BOM); 2012.
[4] Khatib T, Sopian K, Kazem HA. Actual performance and characteristic of a grid connected photovoltaic power system in the tropics: a short term evaluation. Energy Convers Manage 2013;71:115–9.
[5] Golden CB. Solar radiation modeling and measurements for renewable energy applications: data and model quality. NREL, March 2003, 1617 80401-3393.
[6] Kazem HA, Yousif JH, Chaichan MT. Modelling of daily solar energy system prediction using support vector machine for Oman. Int J Appl Eng Res 2016;11 (20):10166–72.
[7] Yousif JH, Kazem HA. Modeling of daily solar energy system prediction using soft computing methods for Oman. Res J Appl Sci, Eng Technol 2016;13 (3):237–44.
[8] Awada A, Pasupuleti J, Khatib T, Kazem HA. Modeling and characterization of a photovoltaic array based on actual performance using cascade-forward back propagation artificial neural network. J Solar Energy Eng: Including Wind Energy Build Energy Conserv 2015;137:4–5.
[9] Mazin H, Kazem HA, Fadhil HA, Aljunid SA, Abdulmajeed QM, Chaichan MT. Linear and nonlinear modeling for solar energy prediction for zone, region and global areas. In: Renewable energy in the service of mankind, vol. II. Springer International Publishing; 2016. p. 21–34.
[10] Argiriou AA et al. A neural network controller for hydronic heating systems of solar buildings. Neural Netw 2012;17:427–40.
[11] Khatib T, Mohamed A, Mahmoud M, Sopian K. Estimating global solar energy using multilayer perception artificial neural network. Int J Energy 2012;6:1.
[12] Dorvloa Atsu SS, Jervaseb JA, Al-Lawati A. Solar radiation estimation using artificial neural networks. Appl Energy 2015;71(4):307–19.
[13] Sulaiman SI, Rahman TKA, Musirin I, Shaari S. An artificial immune-based hybrid multi-layer feedforward neural network for predicting grid-connected photovoltaic system output. Energy Proc 2012;14:260–4.
[14] Mellit A, Kalogirou SA. ANFIS-based modelling for photovoltaic power supply system: a case study. Renew Energy 2011;36(1):250–8.
[15] Khatib T. Optimization of a grid-connected renewable energy system for a case study in Nablus, Palestine. Int J Low Carbon Technol 2014;9(4):311–8.
[16] Modi V, Sukhatme SP. Estimation of daily total and diffuse insolation in India from weather data. Sol Energy 1979;22:407–11.
[17] Page JK. The estimation of monthly Ea values of daily total short wave radiation on vertical and inclined surfaces from sunshine records for latitudes 40°N–40°S. In: Proceedings of the UN conference on new sources of energy, vol. 598(4); 1961. p. 378–90.
[18] Sben Z, Tan E. Simple models of solar radiation data for northwestern part of Turkey. Energy Convers Manage 2001;42:587–98.
[19] Jin Z, Yezheng W, Gang Y. General Formula for estimation of monthly average daily global solar radiation in China. Energy Convers Manage 2005;46:257–68.
[20] Angstrom A. On the computation of global radiation from records of sunshine. Arkiv Geof 1956;2:471–9.
[21] Altunkaynak, M. Ozger. Comments on 'Temporal significant wave height estimation from wind speed by perceptron Kalman filtering 2004;31(0):1245–55.
[22] Jiang Y. Prediction of monthly mean daily diffuse solar radiation using artificial neural networks and comparison with other empirical models. Energy Policy 2008;36:3833–7.
[23] Sahin M, Kaya Y, Uyar M. Comparison of ANN and MLR models for estimating solar radiation in Turkey using NOAA/AVHRR data. Adv Space Res 2013;51:891–904.
[24] McCullock WS, Pitss W. A logical calculus of the ideas immanent in nervous activity. Bull Math Biophys 1943;5:115–33.
[25] Moody T, Darken C. Fast learning in networks of locally tuned processing units. Neural Comput 1989;1:281–94.
[26] Paulescu M, Gravila P, Tulcan-Paulescu E. Fuzzy logic algorithms for atmospheric transmittances of use in solar energy estimation. Energy Convers Manage 2008;49:3691–7.
[27] Jang JSR, ANFIS. Adaptive-network-based fuzzy inference system. IEEE Trans Syst Manage Cybern 1993;23(3):665–85.
[28] Top S, Dilma U, Aslan Z. Study of hourly solar radiation data in Istanbul. Renew Energy 1995;6:171–84.
[29] Kennedy J., Eberhart R. Particle swarm optimization. In Proceedings of the International Conference on Neural Networks 1995; IV:1942–1948.
[30] Kazem HA, Khatib T, Sopian K, Elmenreich W. Performance and feasibility assessment of a 1.4 kW roof top grid-connected photovoltaic power system under desertic weather conditions. Energy Build 2014;82:123–9.
[31] Zadeh LA. Toward a generalized theory of uncertainty (GTU)––an outline. Inf Sci 2005;172(1):1–40.
[32] Yousif JH. Information technology development. Germany: LAP LAMBERT Academic Publishing; 2011. ISBN 9783844316704.
[33] Fekihal MA, Yousif JH. Self-organizing map approach for identifying mental disorders. Int J Comp Appl 2012;45:7.
[34] Uraibi HS, Midi H, Talib AB, Yousif JH. Linear regression model selection based on robust bootstrapping technique. Am J Appl Sci 2009;6(6):1191–8.
[35] Sozen A, Arcaklioglu E, Ozalp M, Caglar N. Forecasting based on neural network approach of solar potential in Turkey. Renew Energy 2005;30:1075–90.
[36] Rehmana S, Mohandes M. Artificial neural network estimation of global solar radiation using air temperature and relative humidity. Energy Policy 2008;36:571–6.
[37] Elminir HK, Azzam YA, Younes FI. Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models. Energy 2007;32:1513–23.
[38] Mohandes M, Rehman S, Halawani TO. Estimation of global solar radiation using artificial neural networks. Renew Energy 1998;14:179–84.
[39] Mubiru J, Banda EJKB. Estimation of monthly average daily global solar irradiation using artificial neural networks. Sol Energy 2008;82:181–7.
[40] Bilgili M, Sahin B, Yasar A. Application of artificial neural networks for the wind speed prediction of target station using reference stations data. Renew Energy 2007;32:2350–60.
[41] Bosch JL, Lopez G, Batlles FJ. Daily solar irradiation estimation over a mountainous area using artificial neural network. Renew Energy 2008;33:1622–8.
[42] Fadare DA. Modeling of solar energy potential in Nigeria using an artificial neural network model. Appl Energy 2009;86:1410–22.
[43] Graditi Ferlito S, Adinolfi G. Comparison of Photovoltaic plant power production prediction methods using a large measured dataset. Renew Energy 2016;90:513–9.
[44] Huang C, Bensoussan A, Edesess M, Tsui KL. Improvement in artificial neural network-based estimation of grid connected photovoltaic power output. Renew Energy 2016;97:838–48.
[45] Chow TT, Hand JW, Strachan PA. Building-integrated PV and thermal applications in a subtropical hotel building. Appl Therm Eng 2003;23:2035–49.
[46] Dixon RK, McGowan E, Onysko G, Scheer RMUS. Energy conservation and efficiency policies: challenges and opportunities. Energy Policy 2010;38:6398–408.
[47] Camargo RL, Zink R, Dorner W, Stoeglehner G. Spatiotemporal modeling of roof-top PV panels for improved technical potential assessment and electricity peak load offsetting at the municipal scale. Comput Environ Urban Syst 2015;52:58–69.
[48] Dubey S, Sarvaiya JN, Seshadri B. Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world. A review. Energy Proc 2013;33:311–21.