IJCSN - International Journal of Computer Science and Network, Volume 9, Issue 2, April 2020
ISSN (Online) : 2277-5420
www.IJCSN.org
Impact Factor: 1.5

# Hybrid Artificial Neural Networks with Boruta Algorithm for Prediction of Global Solar Radiation: Case Study in Saudi Arabia

**[1]Abdulatif Aoihan Alresheedi, [2]Mohammed Abdullah Al-Hagery**

[1] Master degree student, Computer Science Department,
College of Computer, Qassim University
KSA

[2] Computer Science Department, College of Computer, Qassim University
KSA

**Abstract -** Precise predictions of renewable energy sources play a vital role in bringing them into the electric grid. This research presents one of the most powerful machine learning algorithms to forecast the hourly global solar radiance. This study utilizes artificial neural networks (ANN) as the machine learning predictor due to its ability to tackle the nonlinear aspects existing in solar data. The type of the used ANN in this analysis is a multilayer feed-forward back-propagating neural network, denoted (MLFFBPNN). Nevertheless, choosing the ideal set of input variables, known as features, to train the predictive models created, which are typically user-determined, is a continuing, primary obstacle in obtaining high predictive efficiency. Therefore, this study's precise purpose is to forecast global horizontal irradiance by building models of neural networks whose input variables are optimally and systemically chosen by the Boruta Algorithm, a powerful feature selection method. Prediction models were built based on real-world solar data collected for a site known as Buraydah in Saudi Arabia. For the creation of the developed forecasting models, thirteen features of solar data are considered, including month of the year, day of the month, hour of day, air temperature, relative humidity, surface pressure, wind speed at 3 meters, wind direction, peak wind direction at 3 meters, diffuse horizontal irradiance, direct normal irradiance, azimuth angle, and solar zenith angle. The performance of the suggested models was assessed using four of the most common measures of error. the results stress the importance of using feature selection techniques when using computational intelligence models to achieve precise solar radiation predictions.

**Keywords -** *Global horizontal irradiance, Artificial neural networks, Feature selection, Boruta algorithm, Big data, Machine learning.*

## 1. Introduction

The future of the world's energy supply is gradually being influenced by alternative energy sources. This is due to the disadvantages of fossil fuel supplies as well as its negative side effects on climate change and emissions. Renewable energy sources, such as solar power, can achieve the goal of sustainable power supply; yet they are distinguished by high variability in availability and output. Rabid increases in solar power production are one of the negative impacts of rapid weather changes. Yes, the intermittent presence of renewable energy sources that impede the productive use of electrical utilities. Higher penetration of solar energy into the electrical grid produces a more variable output of electricity than with higher wind penetrations [1]. In addition, higher penetration of renewable energy sources leads to the technical operation of the power system and design issues such as system safety, system control, the efficiency of the power factor and optimum power system operation [2]. Additionally, changes to the operation of power systems are required to handle the volatility and instability of solar power, which is due to the high penetration of renewables, including the addition of new ancillary services [3]. Therefore, the high costs of these changes and specifications adversely affect the economic viability of alternative energy sources.

Several possible solutions can handle the engineering problems caused by short term solar power volatility (up to seven days ahead). For example, increasing the amount of demand-side involvement, increasing the rate of coordination to manage allocation, and introducing energy storage systems that are more versatile — but often costlier — [4]. Today, global solar radiation prediction is one of the most powerful and cost-effective ways to integrate more solar power, particularly at current integration rates. Balance authorities can use these predictions to more effectively and safely run electric power systems. Many prediction approaches were adopted

in the literature. Among these, Machine Learning (ML) algorithms are presently the most popular methods of forecasting solar energy, since forecasting is an essential step in the technical and economic design and evaluation of photovoltaic (PV) systems [5]. Several studies have formulated predictions of solar radiation using ML algorithms, after the rise of fast computational technologies as well as systems capable of storing large data sets [6]– [9]. These ML approaches involve artificial neural networks (ANN), vector regression support (SVR), decision tree regression (DTR), and K-Nearest Neighbors (KNN) [10].

ANN is rated as a strong ML tool due to its intrinsic ability to deal with the nonlinear complexity of solar and meteorological data [11]. ANNs have led to a lower mean absolute percentage error (MAPE) in recent studies — but often also a higher $R^2$ — as compared with other regression models [12]. For example, ANN has been used to forecast the Tai power system's electrical load [13]. Furthermore, two ANN networks were established in Salerno, Italy, in an hourly manner to forecast global horizontal irradiance (GHI) and direct normal irradiance (DNI) [14]. ANN models have contributed to good accuracy in the two later studies. In [15], Chiteka and Enweremadu applied ANNs to forecast Zimbabwe's global solar radiation; The latitude and longitude geographic data, and weather forecasting data of the humidity, pressure, clearness index and average temperature were used as the model's input variables. With 10 neurons in the hidden layer, the best ANN structure was obtained, and they suggested a tangent sigmoid transfer method for the input and output layers. The network reached a 99.8 percent determination coefficient ($R^2$).

Feature selection methods can be used for productive pre-processing of data to reduce and prepare data with high dimensional space for ML-based problems. These algorithms are generally classified either into supervised algorithms needing label data or unsupervised algorithms which work without any need for label data. The problem is that there are a significant number of variables influencing the solar radiation concentration. Therefore, the prediction accuracy of established forecasting models could be positively impacted by extracting redundant features by using feature selection algorithms. In this analysis, a sophisticated embedded feature selection method known as the Boruta algorithm (BA) is applied to pick the most significant features among 13 features available.

The main contribution of this study lies in the fact that the feature selection method used for this research, i.e. BA, has still not been introduced to this problem. This actually adds significant value to this article. Since this paper's study requires large data sets, we adopt the powerful machine learning algorithm of ANN as a computing system tool. The kind of ANN, known as multilayer-feed-forward-back-propagation-based neural networks (MLFFBPNN), is utilized during this research. Shortly, this article presents a new, smart, hybrid system consisting of the ANN algorithm for performing training and testing processes, and BA for determining the most essential features to be incorporated into the ANN model. Consequently, this article recommends 3 ML-based models to forecast GHI for the site of interest. One can summarize the suggested models as follows:

- GHI forecasting model based on the ANN algorithm with a set of all features.
- GHI forecasting model based on the ANN algorithm with a set of the best eight features suggested by BA.
- GHI forecasting model based on the ANN algorithm with a set of the best five features suggested by BA.

The remainder of this essay is structured in a sectional way. First of all, section 2 shines a light on the architecture of the hybrid prediction models developed including the data used to train them as well as the basics of the ML algorithms used. A description of the feature selection method used in this article is provided also in Section 2. Section 3 addresses error measures that are used to determine the accuracy of the suggested models. The recording of the findings of the study and their discussion can be found in section 4. Eventually, section 5 reviews the conclusions of the article and provides some suggestions for future work.

## 2. Methodology

This section basically shines a light on the general context of the research as well as the data used in the analysis. What's more, the fundamentals of ANN are clarified tremendously.

### 2.1 Proposed network

Fig.1 displays the conceptual structure of the planned analysis, which is outlined below as:

- Data pre-processing activities of all data sets used throughout the analysis shall be performed prior to the training validation and testing of the data-based models proposed.
- MLFFBPNN model is developed which is trained and validated using the all-feature set.
- The MLFFBPNN model, trained and validated using the most appropriate eight-function set defined by the BA, is also constructed.
- In addition, the MLFFBPNN model, trained and validated using BA's most significant five-function set, is also developed.

- The GHI values predicted by the three different models produced are performed during the testing process.
- To assess the accuracy of the prediction models built in this analysis, a collection of four assessment metrics compares the actual and the forecast values of the GHI.

## 2.2 Data collection

Massively large observed solar and meteorological data sets were collected from King Abdullah City for Atomic and Renewable Energy, denoted K.A.CARE, to create the proposed forecasting models presented in this study. While we wish to create more location-specific models, the Buraydah site in Saudi Arabia is chosen as the main location of the study, as shown in Fig.2. Henceforth, this site of interest will be recognized and described as: Buraydah in this article. The GHI observations under consideration are collected for the interest location of Buraydah in 1-hour time resolutions for the timeframe from March 1, 2013, to June 30, 2017. The entire data *(X)* is divided into three sub-sets, which include: the training dataset, $x_{training}$, the cross-validation dataset, $x_{cross\text{-}validation}$, and the testing dataset, $x_{testing}$, such that $X=( x_{training} \cup x_{cross\text{-}validation} \cup x_{testing} )$. The ratio of the training, cross-validation, and testing datasets in this analysis is 6:2:2, respectively.

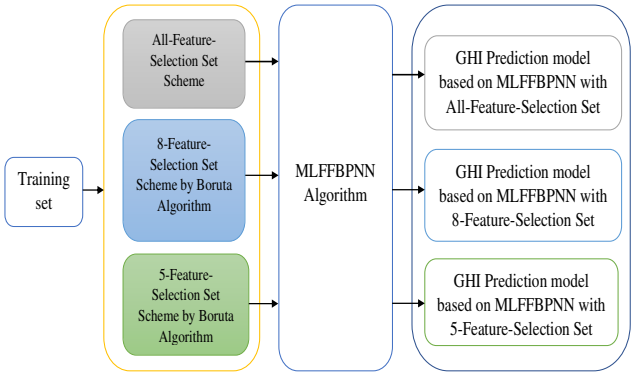


Fig. 1 The schematic illustration of the proposed data-driven models for global horizontal irradiance.

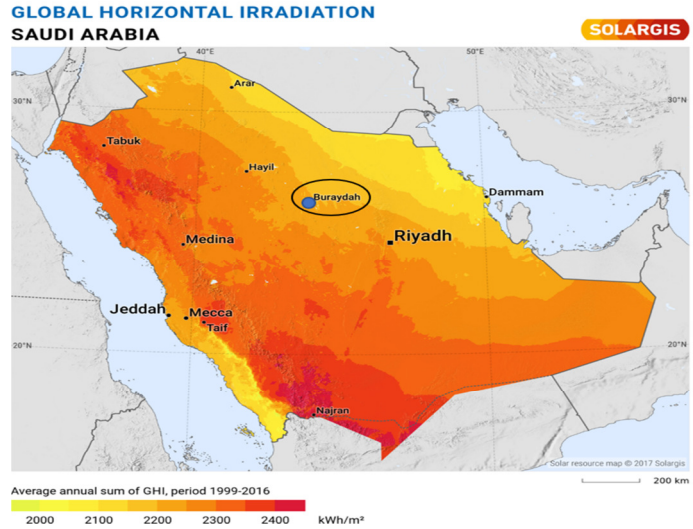


Fig. 2 The oval shows the present study area located in Buraydah province, Saudi Arabia, showing the average annual sum of GHI over 1999-2000.

The parameters included in this study are 13 independent variables: month of the year (M), day of the month (D), hour of the day (H), air temperature (T) in (℃), relative humidity (RH) in (%), surface pressure (P) in (hPa), wind speed at 3 meters (WS) in (m/s), Wind Direction (WD) at 3 meters in (°N), Peak Wind Direction (PWD) at 3 meters in (°N), diffuse horizontal irradiance (DHI) in (Wh/m$^2$), direct normal irradiance (DNI) in (Wh/m$^2$), azimuth angle (AA) in (), and solar zenith angle (SZA). ). Global solar irradiance GHI in (Wh/m$^2$), as a target variable. Table 1 summarizes the input variables to the prediction models. The predictive models can be described, as shown in Eq. (1) in the context of the predictors used:

$$GHI_{predicted} = f\,(M, D, H, T, RH, P, WS, WD, PWD, DHI, DNI, AA, SZA) \qquad (1)$$

Table 1: Input variables to the proposed prediction models

| *Input variable* | *Input variable abbreviation* | *Input variable explanation* | **Input variable unit** |
|---|---|---|---|
| $x_1^{(i)}$ | M | month of the year | Month |
| $x_2^{(i)}$ | D | day of the month | day |
| $x_3^{(i)}$ | H | hour of the day | hour |
| $x_4^{(i)}$ | T | air temperature | ℃ |
| $x_5^{(i)}$ | RH | relative humidity | % |
| $x_6^{(i)}$ | P | surface pressure | hPa |
| $x_7^{(i)}$ | WS | wind speed at 3 meters | m/s |
| $x_8^{(i)}$ | WD | Wind Direction | °N |
| $x_9^{(i)}$ | PWD | peak wind direction at 3 meters | °N |
| $x_{10}^{(i)}$ | DHI | diffuse horizontal irradiance | Wh/m2 |
| $x_{11}^{(i)}$ | DNI | direct normal irradiance | Wh/m2 |
| $x_{12}^{(i)}$ | AA | azimuth angle | Â° |
| $x_{13}^{(i)}$ | SZA | solar zenith angle | Â° |

## 2.3 Input Data Normalization

The scaling of input data, also known via normalization, is a very important practical functionality when applying ANNs. The purpose of this practical application is primarily to avoid the possibility of dominating attributes with higher numerical values on those attributes with lower ones. In this research, using Eq. (2), every attribute is normalized linearly to the range [0, 1] as below:

$$x_i^n = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Where,

- $x_i$ is the actual value of the feature vector;
- $x_{min}$ and $x_{max}$ are the minimum and the maximum values corresponding to the actual dataset.

## 2.4 Multilayer Feed-forward Back Propagating Neural Networks (MLFFBPNN)

Artificial Intelligence (AI) approaches can capture non-linear interactions between the independent and dependent variables. ANN is one of the strong non-linear forecast algorithms used here. Being ANNs is capable of modeling non-linear processes without having to presume that one of the main advantages of this approach is the relationship type between the input and output variables. The ANN form used here, Fig.3, is an MLFFBPNN, also called a multilayer perception [16]. The back-propagation algorithm (BP) is chosen here in this study to perform the MLFFBPNN training process since it is one of the most popular ANN algorithms.

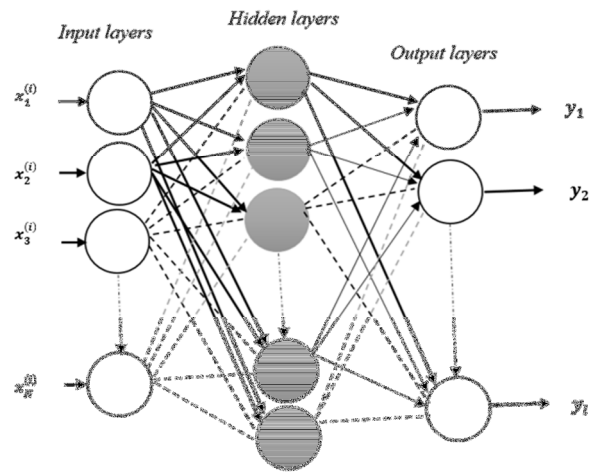


Fig. 3 Multilayer-Feed-forward Back Propagating Neural Networks (MLFFBPNN).

As Fig.3 depicts, the usual architecture of ANNs formed with three main layers. First, the input layer, $[x_1, x_2, \dots, x_N]^T$, is composed of an N-dimensional input vector. After that, the hidden layer, $[h_1, h_2, \dots, h_M]^T$, which includes a nonlinear activation function known as the activation function. Finally, the output layer, $[y_1, y_2, \dots, y_L]^T$, which contains a linear function. Inside hidden layers, the output values of the neurons, which represent the basic component of any ANN. Mathematically, the ANN model can be summarized by Eq. (3) as bellow:

$$y(x) = f(b + \sum_{i=1}^{n} w_i(p).x_i(p)) \quad (3)$$

In which,

- $x_i(p)$ represents the input variable;
- $y(x)$ is the output value of the target;
- n is the optimal number of hidden neurons;

- $w_i(p)$ is the weight connecting the i[th] neuron in the input layer to each incoming connection;
- b represents the bias term;
- $f(.)$ represents the hidden activation function.

### 2.5 Feature Selection Algorithm Boruta Algorithm (BA)

BA is a wrapper tool established on the basis of the algorithm of random forest regression and executed by the randomForest package in R [17]. This paper used the BA regression version. The phases the BA is executed in are as follows:

• In order to expand the information system, add copies of all features. Even when the number of features in the original datasets is less than 5, the information system should always be extended by at least 5 shadow features.
• Shuffle the added functions to remove their association with the output variable.
• To collect the calculated Z scores run on the extended information system, a random forest classifier.
• Figure the maximum Z score among shadow features (MZSF), and give any better-scored feature a hit afterward than MZSF.
• For each feature of undetermined importance, a two-sided equality test is performed with the MZSF.
• List the features that are significantly lower than MZSF as ' unimportant ' and exclude them from the information system permanently.
• Consider as 'important' the features of significantly higher value than MZSF.
• Erase all features of the shadow.
• Repeat the process until the value has been allocated for all the features, or the algorithm has reached previously set random forest runs.

Finally, A hybrid strategy for short-term global solar radiation forecasting based on ANN and BA is implemented, which is used to pick the optimal set of features to be inputted into the ANN algorithm. The entire structure of the hybrid system formed can be shown in Fig. 4.

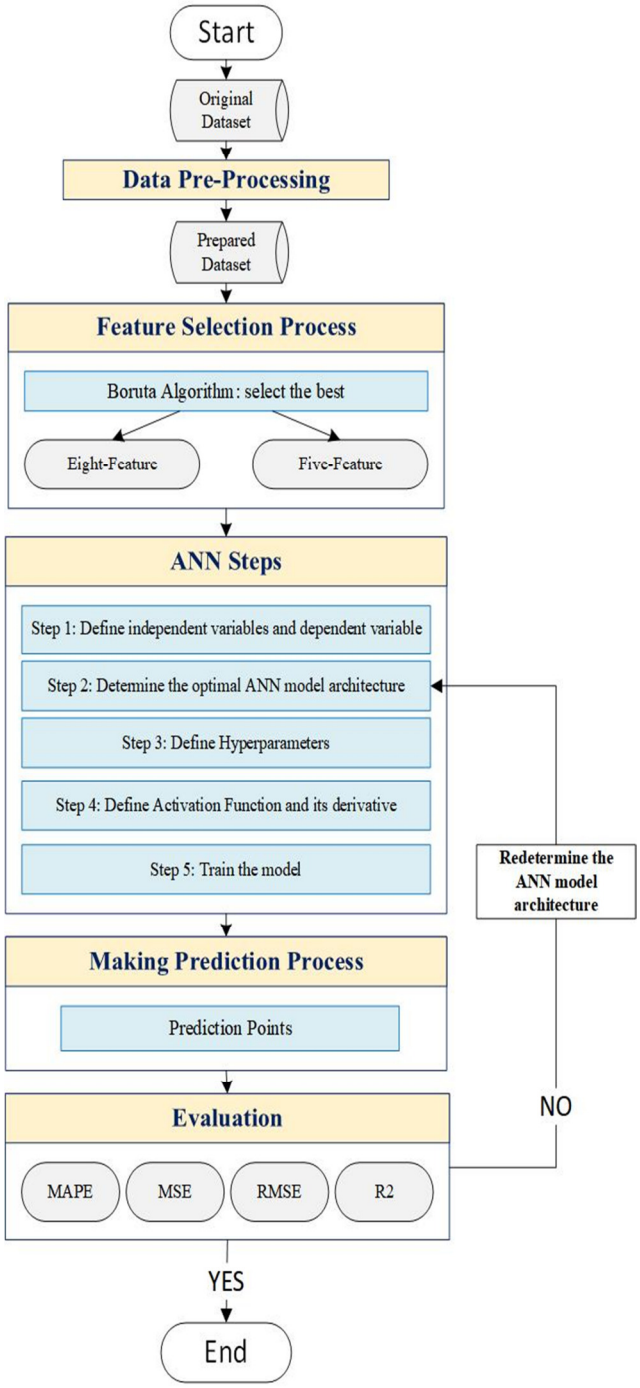


Fig. 4 The flowchart of the proposed global horizontal irradiance forecasting model used in this study.

## 3. Performance Evaluation Metrics

The predictive output accuracy of the built models can be measured using several statistical indicators. This analysis primarily takes into account four metrics: mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), and fitness

goodness ($R^2$) [8]. These measurements are expressed in mathematical terms by Eqs. [ 4]–[7 ], as below, respectively:

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|f_i - y_i|}{y_i} \times 100\% \quad (4)$$

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(f_i - y_i)^2 \quad (5)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(f_i - y_i)^2} \quad (6)$$

$$R^2 = \frac{\sum_{i=1}^{N}(f_i - \tilde{y})^2}{\sum_{i=1}^{N}(y_i - \tilde{y})^2} \quad (7)$$

Where

- N represents the number of data points involved in the analysis;
- $y_i$ is the observed value of the target;
- $f_i$ is the predicted value of the target;
- $\tilde{y}$ is the mean of the observed value of the target $y_i$.

## 4. Results and Discussion

The forecast models were used to estimate GHI at Saudi Arabia's site of interest in Buraydah Provence. In this analysis the models used for prediction were as follows:

(1) MLFFBPNN based on the All-Feature model;
(2) MLFFBPNN based on the Eight-Feature model; and
(3) MLFFBPNN based on the Five-Feature model.

The MLFFBPNN has been implemented using MATLAB R2019a in this analysis. The data were first cleaned and then normalized to increase the efficiency of the forecasting algorithms as well as feature selection algorithms. To build, the so-called Eight-Feature and Five-Feature forecasting models, a collection of eight and five features were defined out of the available 13 features. The selection of these features was based on the feature selection algorithm, Boruta Algorithm (BA). Eight and five features were chosen to show the success with a different number of features of different forecasting models. Fig.5 orders the features according to their significance for our outcome (GHI) at the site of Buraydah.

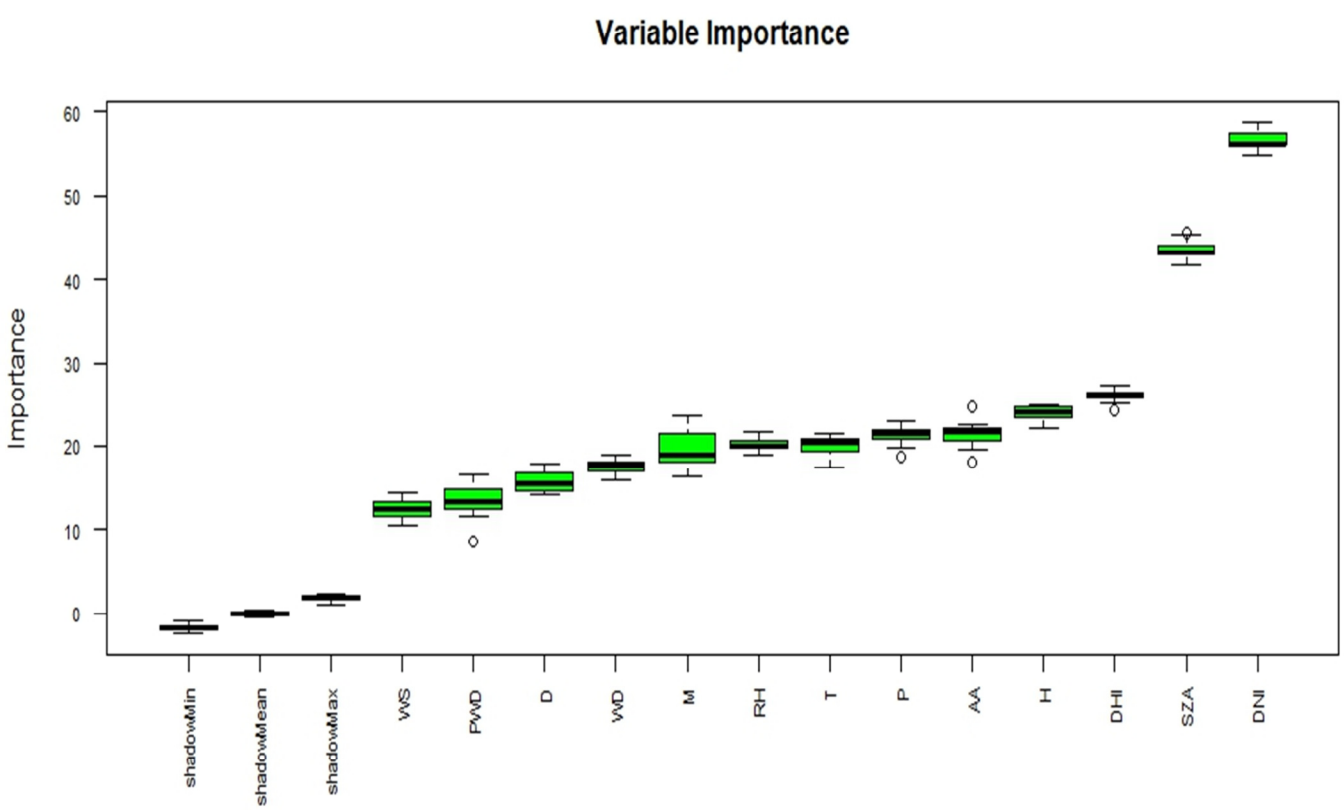


Fig. 5 Variable importance for weather data for Buraydah using the Boruta algorithm.

Fig.5 shows that the Eight-Feature model is developed using the following features: DNI, Solar Zenith Angle (SZA), DHI, Hour (H), Azimuth Angle (AA), Pressure (P), Temperature (T), and finally our target variable GHI. For the Five-Feature model, the forecasting model built based on the following input variables: DNI, Solar Zenith Angle (SZA), DHI, Hour (H), and finally our target variable GHI.

Training networks for all three models used MLP with a BP, while the training function was the Levenberg-Marquardt method. One input layer, one hidden layer, and one output layer are used. For all-feature, eight-feature, and five-feature models, the hidden layer in the ANN network was designed with 14, 8, and 5 neurons (nodes), respectively. The training and testing dataset input and output are identical to all models.

The results of the All-Feature model were compared with the results based on both Eight-Feature and Five-Feature models in terms of their predictability at the location of Buraydah. To this end, the test models ' accuracy outputs were evaluated on the basis of the following metrics: MAPE, MSE, RMSE, and $R^2$. The forecast results of all of Buraydah's predictive models are compared in Table 2.

Table 2: Results of hourly forecasted GHI at Buraydah

| | MAPE (%) | MSE Watt/m$^2$ | RMSE Watt/m$^2$ | $R^2$ |
|---|---|---|---|---|
| All-Feature | 13.701 | 994.3696 | 31.533 | 0.9912 |
| Eight-Feature | 12.202 | 939.868 | 30.657 | 0.9916 |
| Five-Feature | 9.693 | 913.843 | 30.229 | 0.9918 |

The Five-Feature model can be considered as the best model followed by the Eight-Feature and All-Feature models, respectively, according to Table 2 which presents Buraydah results. Buraydah's forecasting models with only five and eight features show how important it is to use a feature selection approach, and how adding more features will lead to overfitting forecasting model. In the site of the interest, the MAPE value was determined to be 13.7052% with All-Feature (12.2033% with Eight-Feature and 9.6235% with Five-Feature). While MSE values of All-Feature model are 991.3696 Watt/m$^2$ (930.8685 Watt/m$^2$ with Eight-Feature and 912.8432 Watt/m$^2$ with Five-Feature), the RMSE values of All-Feature model are 31.5336 Watt/m$^2$ (30.6572 Watt/m$^2$ with Eight-Feature and 30.2298 Watt/m$^2$ with Five-Feature), respectively. Buraydah's forecast models correlation scores were found to be 0.991243, 0.991677, and 0.991836 for All-Feature, Eight-Feature, and Five-Feature, respectively.

For further visualization, the measured GHI values are plotted against the output of the three forecasting models at Buraydah (see Fig.6). Fig.6 reveals how the Five-Feature model performs better in comparison followed by Eight-Feature and All-Feature models, respectively.

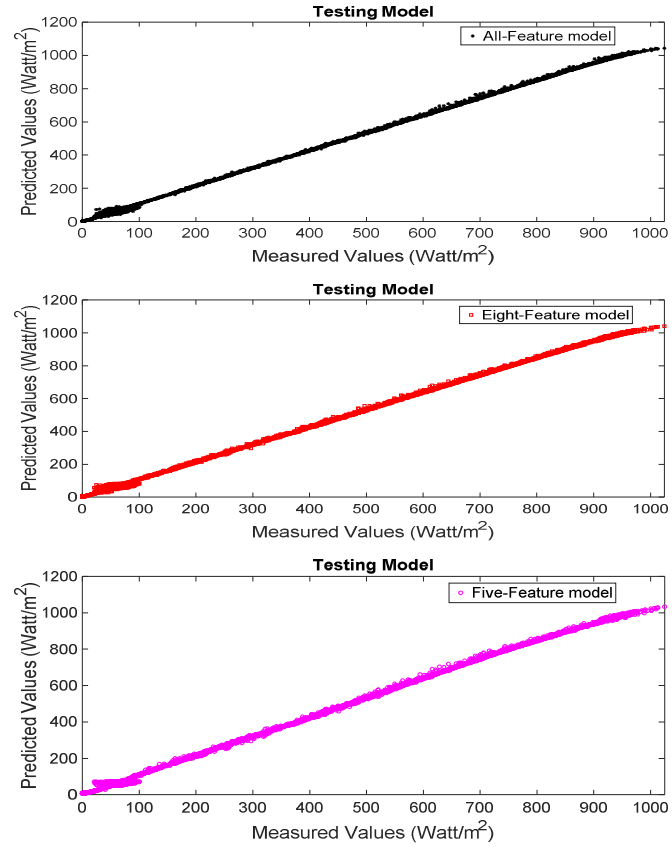


Fig. 6 Measured vs. predicted values of GHI for Buraydah, All-Feature model (top), Eight-Feature model (middle), and Five-Feature model (bottom).

The results of all-Feature, Eight-Feature, and Five-Feature models are plotted in Fig.7 for twenty random hours together with measurements of GHI values. Fig.7 confirms the Five-Feature model's capacity to match the GHI values calculated at the location of interest.

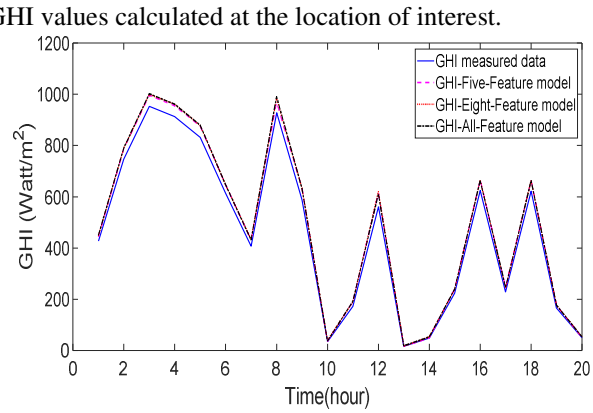


Fig. 7 The three forecasted GHI models values vs measured GHI values at Buraydah.

## 5. Conclusion And Future Work

This paper discusses the use of an advanced embedded feature selection algorithm and artificial neural networks to forecast the hourly solar radiation at the site of Buraydah in Saudi Arabia. Data were obtained from one of the monitoring stations in Saudi Arabia to analyze the performance of the established models in prediction. The five and eight most significant variables among a wide variety of weather variables that might influence solar radiation in the future were defined optimally and systematically using a modern selection technique called Boruta algorithm. In addition, and mainly for purposes of comparison, an all-feature model was developed to determine the advantages of using a selection method of features.

At the site of interest, and according to the results of the feature selection algorithm, the use of the five-feature model resulted in the best predictive performance compared to the predictive values of the models that used a greater number of eight or all features. While the changes in MAPE were found to be 9.6 % when using the Five-Feature model, they are found to be 12.2 % and 13.7 % respectively for the Eight-Feature and All-Feature models. Therefore, using feature selection techniques, the larger interdependent variables applicable to the hourly global horizontal irradiance forecast can be effectively exploited without compromising predictive performance. Consequently, the results stress the importance of using feature selection techniques when using computational intelligence models to achieve precise solar radiation predictions. Future research based on this paper may be expanded to include evaluating the influence of other feature selection techniques in selecting the most important features for forecasting global solar radiation.

## Acknowledgments

The first researcher, Abdulatif Alresheedi, would like to thank Qassim University, Saudi Arabia, for academically supporting this study. Furthermore, the gratitude goes to King Abdullah City for Atomic and Renewable Energy (K.A.CARE) for supplying us with the datasets used in this study.

## References

[1] D. Lew et al., "Sub-Hourly Impacts of High Solar Penetrations in the Western United States," To be Present. 2nd Annu. Int. Work. Integr. Sol. Power into Power Syst. Conf. Lisbon, Port., pp. 1–9, 2012.

[2] E. M. Sandhu and T. Thakur, "Issues, Challenges, Causes, Impacts and Utilization of Renewable Energy Sources -Grid Integration," J. Eng. Res. Appl., vol. 4, no. 1, pp. 636–643, 2014.

[3] I. J. Perez-Arriage, "Managing large-scale penetration of intermittent renewables," Anesthesiology, vol. 115, no. 3, p. A13, 2011.

[4] A. Tuohy et al., "Solar Forecasting: Methods, Challenges, and Performance," IEEE Power Energy Mag., vol. 13, no. 6, pp. 50–59, 2015.

[5] S. Shamshirband, M. Danesh, M. Hossain, S. Mekhilef, and L. Olatomiwa, "Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems," J. Clean. Prod., vol. 167, pp. 395–405, 2017.

[6] J. Polo, A. Navarro, M. Cony, L. Martín, R. Marchante, and L. F. Zarzalejo, "Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning," Sol. Energy, vol. 84, no. 10, pp. 1772–1781, 2010.

[7] A. Alzahrani, J. W. Kimball, and C. Dagli, "Predicting solar irradiance using time series neural networks," Procedia Comput. Sci., vol. 36, no. C, pp. 623–628, 2014.

[8] A. Sharma and A. Kakkar, "Forecasting daily global solar irradiance generation using machine learning," Renew. Sustain. Energy Rev., vol. 82, no. May 2017, pp. 2254–2269, 2018.

[9] I. A. Ibrahim and T. Khatib, "A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm," Energy Convers. Manag., vol. 138, pp. 413–425, 2017.

[10] S. Ghimire, R. C. Deo, N. J. Downs, and N. Raj, "Self-adaptive differential evolutionary extreme learning machines for long-term solar radiation prediction with remotely-sensed MODIS satellite and Reanalysis atmospheric products in solar-rich cities," Remote Sens. Environ., vol. 212, no. November 2017, pp. 176–198, 2018.

[11] S. Haykin, Neural Networks: A Comprehensive Foundation, 2nd ed. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1998.

[12] R. Kumar, R. K. Aggarwal, and J. D. Sharma, "Comparison of regression and artificial neural network models for estimation of global solar radiations," Renew. Sustain. Energy Rev., vol. 52, pp. 1294–1299, 2015.

[13] W.-Y. Chang, "Short-Term Load Forecasting Using Radial Basis Function Neural Network," J. Comput. Commun., vol. 03, no. 11, pp. 40–45, 2015.

[14] C. Renno, F. Petito, and A. Gatto, "Artificial neural network models for predicting the solar radiation as input of a concentrating photovoltaic system," Energy Convers. Manag., vol. 106, pp. 999–1012, 2015.

[15] Chiteka, K., Enweremadu, C.C., 2016. Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks. J. Clean. Prod. 135, 701e711.

[16] F. Rosenblatt, "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain," Psy

chol. Rev., vol. 65, no. 6, pp. 386–408, 1956.

[17] Kursa, Miron B., and Witold R. Rudnicki. "Feature Selection with the Boruta package." J Stat Softw 36.11 (2010): 1-13.

**Authors –**

**Abdulatif Alresheedi:** Currently is a Master's student at Qassim University in Computer college, Computer Science department. Alresheedi received his Bachelor's degree in Computer Science in 2004. From 2004 until now, I joined the public sector as the manager of Information Technology Management. My research interests lie in data mining, machine learning algorithms, advance optimization techniques, and big data.

**Mohammed Abdullah Al-Hagery**: received his B.Sc in Computer Science from the University of Technology in Baghdad Iraq-1994. He got his MSc in Computer Science from the University of Science and Technology Yemen-1998. AlHagery finished his Ph.D. in Computer Science and in Information Technology, (Software Engineering) from the Faculty of Computer Science and IT, University of Putra Malaysia (UPM), 2004. He was head of the Computer Science Department at the College of Science and Engineering, USTY, Sana'a from 2004 to 2007. From 2007 to this date, he is a staff member at the Faculty of Computer, Department of Computer Science, Qassim University in KSA. He published more than 15 papers in international journals. Dr. Al-Hagery was appointed head of the Research Centre at the Computer College, Qassim University, KSA from September 2012 to October 2018.