An Artificial Neural Network Model for Estimating Daily Solar Radiation in Northwest Nigeria

1Salisu Aliyu, 2Aminu S. Zakari, 3Muhammad Isma’il and 4Mohammed A. Ahmed
1Department of Computer Science, Ahmadu Bello University, Zaria, Nigeria
2Department of Computer Science, Kaduna Polytechnic, Kaduna, Nigeria
3Department of Geography and Environmental Studies, Ahmadu Bello University, Zaria, Nigeria
aliyu salisu@abu.edu.ng | [aminusalihi mahmed]@kadunapolytechnic.edu.ng | mgeogjameel@yahoo.com

Received: 05-AUG-2020; Reviewed: 04-SEP-2020; Accepted: 10-SEP-2020
http://dx.doi.org/10.46792/fuoyejet.v5i2.563

Abstract - Solar energy has attracted enormous attention as it plays an essential role in meeting the ever growing sustainable and environmentally friendly energy demand of the world. Due to the high cost of calibration and maintenance of designated measuring instruments, solar radiation data are limited not only in Nigeria but in most parts of the world. The optimal design of solar energy systems requires accurate estimation of solar radiation. Existing studies are focused on the analysis of monthly or annual solar radiation with less attention paid to the determination of daily solar radiation. Estimating daily solar radiation envisages high quality and performance of solar systems. In this paper, an Artificial Neural Network data mining model is proposed for estimating the daily solar radiation in Kano, Kaduna and Katsina, North West Nigeria. Daily Solar radiation data for 21 years collected from the Nigerian Meteorological Agency were used as training/testing data while developing the model. Two statistical indicators: coefficient of determination (R²) and the root mean square error (RMSE) were used to evaluate the model. An RMSE of 0.47 and 0.48 was obtained for the training and testing dataset respectively, while an R² of 0.78 was obtained for both the training and testing dataset. The overall results showed that artificial neural network exhibits excellent performance for the estimation of daily solar radiation.

Keywords - Artificial Neural Network, Data mining, Solar Radiation

1 INTRODUCTION

The depletion of fossil energy sources, photochemical pollution, acid rain, emission of CO₂ and other hazardous greenhouse gases are some of the disadvantages of non-renewable energy. Renewable energy resources have gained significant importance in the 21st century as researchers are striving hard to make a pollution free environment and sustainable energy provision (Inayat & Raza, 2019). Renewable energy is available through different natural resources such as wind, solar, geothermal etc. Solar energy has attracted enormous attention as it plays an essential role in meeting the ever growing sustainable and environmentally friendly energy demand of the world. Solar radiation data is indispensable for developing and accessing solar energy utilization technologies. Although, practically measured solar radiation data is more accurate, however it is not readily available mainly due to the initial investment and maintenance cost of measuring instruments (Notton et al., 2019). The difficulties and uncertainty involved in measurement and collection of solar radiation data have resulted in development of many models and algorithms for its estimation. Hence, optimal design of solar energy systems now relies on the accurate estimation of solar radiation data.

Data Mining is the process of selecting, exploring, and modelling of large amount of data to discover unknown patterns or relationships which provides a clear and useful result to data analyst. Data mining has been applied with success to different fields of human endeavours such as energy and health (Aliyu, Musa & Jauro, 2018) for building predictive, descriptive, forecasting or estimating models. In the past decade, several data mining algorithms have evolved such as Support Vector Machine (SVM), K-Nearest Neighbour, Naïve Bayes, Artificial Neural Network (ANN), Decision Trees etc. that seeks to foretell some response of interest.

Several studies have shown the use of data mining techniques in estimating solar radiation with ANN being the most popular due to its highly non-linear modelling characteristics. According to the literature, Şenkal and Kuleli (2009) studied ANN model for 12 cities in Turkey. The inputs were latitude, longitude, altitude, month, mean diffuse radiation and mean beam radiation. 5 months of data from 9 stations were used to train the ANN, while 5 months of data from other 3 stations were used for testing. The study also proposed a satellite-based method to estimate the monthly average daily radiation. The RMSE of ANN and satellite-based method in training stage were 2.32 and 2.75 MJ/(m² day) respectively, and in testing stage were 3.94 and 5.37 MJ/(m² day) respectively.

Şenkal (2010) built an ANN model using latitude, longitude, altitude, two types of surface emissivity and land surface temperature as inputs. One-year data from 10 stations was used to train the ANN and one year of data from other 9 stations was used for testing. The RMSE in training and testing stage were 0.16 and 0.32 MJ/(m² day), respectively. Qin et al. (2011) built a feed-forward ANN with single hidden layer, which used 6 remote sensing products as inputs, including difference of land surface temperature between daytime and night time, mean land surface temperature, monthly precipitation, enhanced vegetation index, number of days and air pressure ratio. They selected Levenberg-Marquardt algorithm coupled with Bayesian regulation to train the ANN and used hyperbolic tangent sigmoid function as the transfer function in hidden layer. There were 12 neurons in the hidden layer. 7-years data from 22 stations on Tibetan Plateau were used to train the ANN and 5 years of data from other 12 stations was used for testing. The RMSE in testing stage was 1.40 MJ/(m² day) and the relative RMSE is 8.47%.
Ozgoren, Bilgili and Sahin (2012) developed a feed-forward ANN for Turkey. The method of stepwise multi-non-linear regression was applied to determine the most suitable independent variables for the input layer. Ten variables were selected, including sunshine duration, month of the year, cloudiness, soil temperature, mean atmospheric temperature, altitude, wind speed, maximum atmospheric temperature, minimum atmospheric temperature, and latitude. There were 10 neurons in the hidden layer. Levenberg-Marquardt optimization algorithm was used to train the ANN. A 7-year data from 27 stations was used for training while 7-year data from other 4 stations was used for testing. The Mean Absolute Percentage Error (MAPE) in training and testing stage were 4.29% and 5.34% respectively.

Mohandes (2012) used particle swarm optimization to train the ANN developed for Saudi Arabia. The inputs were month of the year, latitude, longitude, altitude, and sunshine duration. However, the estimation was for long-term monthly average daily global solar radiation. The data from 31 stations were used to train the ANN while the data from other 10 stations were used for testing. The average MAPE is 8.85%. Yadav, Malik and Chandel (2015) applied the software of Waikato Environment for Knowledge Analysis to find the most influencing input parameters for estimation. They identified the average temperature, maximum temperature, minimum temperature, altitude, and sunshine duration as the most relevant input variables, while latitude and longitude were the least influencing variables. However, the estimation was also for long-term monthly average daily global solar radiation. An ANN with 10 neurons in hidden layer achieved the best performance, and the maximum MAPE was 6.89%.

Jiang et al. (2015) added a hard-ridge penalty to radial basis function (RBF) to reduce the number of nodes in the hidden layer. They applied Pearson correlation coefficients and Apriori association rules to select the relevant input variables. 12 parameters were selected, including average total opaque sky cover, opaque sky cover, precipitation, broadband aerosol optical depth, maximum temperature, minimum temperature, average temperature, daylight temperature, relative humidity, heating degree days, cooling degree days and wind speed. 11years of data from 4 sites in the United States were used to train the ANN while 2 years of data were used for testing. The Cuckoo Search hard-ridge-RBF model achieved the best performance, and the MAPE of the four sites ranged from 5.44% to 7.64% and the RMSE ranged from 1.01 to 1.42 MJ/(m²-day).

Olatomiwa et al. (2015) studied adaptive neuro-fuzzy inference system (ANFIS) for Iseyin, Nigeria. ANFIS was a hybrid intelligent system that combines the learning power of ANN with fuzzy logic. The inputs were maximum temperature, minimum temperature, and sunshine duration. 15 years of data were used to train the model while 6 years of data were used for testing. The RMSE in training and testing stages were 1.09 and 1.76 MJ/(m²/day), respectively. Bou-Rabee et al. (2017) developed a forecasting model based on Artificial Neural Networks to forecast average daily solar radiation of five different Kuwait cities. A feed-forward ANN model that consist of an input layer, an output layer and one hidden layer with 10 neurons was developed. Kazem, Yousif and Chaichan (2016) developed a prediction model of daily solar energy system using Support Vector Machine for Oman. Solar radiation and ambient temperature were used as input that resulted in a MSE of 0.0263 and R² of 0.0774. Naveen (2018) developed six ANN architecture with back propagation algorithm with different input combination and used to predict the monthly mean daily global solar radiation with an RMSE of 3.96.

From the above review, it will be observed that there are two major categories of estimating solar radiation: Monthly daily average solar radiation (MDASR) and daily solar radiation (DSR). MDASR also referred to as a long-term daily radiation suffers from irregularity in the process of data reduction that can affect the result. This is because the data used cannot reflect the condition of an actual day and if used means that the solar radiation condition does not change between adjacent years. The daily solar radiation on the other hand and the hourly radiation data has proven to be much more useful for engineers but unfortunately most daily estimate from the literature used a lot of parameters that may not be readily available in this part of the world. Since sunshine duration has been discovered to be the most influential parameter, the present study hence focuses on the estimation of the daily solar radiation from measured sunshine duration using ANN. It is also worth noting that in the context of Northwest Nigeria, data on other variables that influences solar radiation as attainable in other literature are unavailable largely due to inadequate or absence of measuring instruments. As a result of these, the daily solar radiation data collected was used as a time series data, an approach not considered in most of the literature. The main contribution of the present study is the demonstration of the viability of building an artificial neural network model for a single time series data input.

The rest of this paper is organized as follows. Section 2 gives a brief description of the study area and data collection for the study. Section 3 provides details on the methodology used for model development and strategies used for evaluating the developed model. Section 4 presents results obtained and its interpretation. Section 5 provides concluding remarks on the present study.

2 Study Area
The study area comprises of Katsina, Kano, and Kaduna states in Northwest Nigeria, covering an area of 90,376Km². The area is located between Latitudes 90 00N – 130 30 North of the Equator, and Longitudes 60 00E – 90 30 east of Greenwich Meridian as shown in Fig. 1. The area is bordered by Niger Republic to the north, Bauchi State to east, Abuja to the south and Zamfara State to the west. The study area is located on the “High Plains of Northern Nigeria” which lies on altitude of between 450 to 745m above sea level (Udo, 1970). The climate of the area is described based on Koppen’s classification, characterized by alternating wet and dry seasons. The dry season is characterized by a dust laden wind from the Sahara Desert, known as Harmattan, brought by the
tropical continental air mass, while the rainy season is heavily influenced by the tropical maritime from the Atlantic Ocean, occurring between April/May and September/October (Eludoyin, 2011). The temperature in the area interchanges between hot and cold seasons with the occurrence of Harmattan (a very cold season) between November and February. The Harmattan greatly reduces the amount of solar radiation received in the area. The average minimum temperature in the area is $21.4 \pm 3.43^\circ C$. The mean maximum temperature in the area is $32.8 \pm 3.37^\circ C$ (Eludoyin, 2011).

The area is characterized by leached ferruginous tropical soil developed on deeply weathered Precambrian Basement Complex rock. The projected population of the area in 2016 was about twenty-seven million.

The study used ground-measured solar radiation data obtained from the Nigerian Meteorological Agency (NIMET) stations. The NIMET stations use the same calibrations for collecting solar radiation data. The stations use solarimeter to measure solar radiation in watt per meter square (Wm$^2$). For this study, daily average solar radiation data for a period of 21 years were provided by NIMET in millijoule per meter square per day (MJ/m$^2$/day). The data were converted into kilowatt per hour per meter square per day (kWh/m$^2$/day) using the International Energy Agency (IEA) General Converter for Energy. This is because solar radiation values are generally expressed in kWh/m$^2$/day. The measured values of daily sunshine radiation are shown in Fig. 2.

### 3 Methodology

#### 3.1 Artificial Neural Network

Artificial Neural Network (ANN) is a numerical modelling technique that is inspired by biological neural system and is capable of processing non-linear relationship, data sorting, pattern detection, optimization, clustering, and simulation. An ANN model is usually made up of three layers: Input layer, Hidden layer and Output layer as depicted in Fig. 3.

Each neuron in the hidden layer is guided by the computational rule expressed mathematically as:

$$ y = g \left( \sum_{i=1}^{n} w_i x_i \right) $$

Where:
- $x_i$ = input parameter
- $w_i$ = weight corresponding to each input parameter
- $g(.)$ = transfer function which can be either linear, sigmoid, or hyperbolic

For a multilayer perceptron network with more than one neuron as depicted also in Fig. 3, the output vector $Y$ is expressed as:

$$ Y = f(X, W) $$

Where:
- $X$ = input vector
- $W$ = weight vector
- $f(.)$ = functional relationship between input and output vector.

To mine and discover pattern from a given data set, ANN works with two splits of the dataset known as training dataset and testing dataset. With the training dataset, ANN keeps adjusting weight corresponding to each input parameter to minimize difference between the computed outputs and observations (original output values) which describes the learning process. This process continues until an acceptable weight is determined which will then be used to estimate the output of a testing dataset.

It is worth noting that any modification on ANN configuration variables such as number of layers, number of neurons, training algorithms, transfer function etc. creates a new model. In this work, a sequential model was developed with an input dimension of 1 to the first layer. This is because we are considering our input data as a
time series data as opposed to previous works. The Rectified Linear Unit (Relu) activation function, a loss function, mean squared error and the Adam optimizer were used. Also, the model was designed to stop training when a monitored loss has stopped improving. A value patience=2, indicating the number of epochs with no improvement after which training will be stopped was set and the ANN was trained for 100 epochs and a batch size of 1 was used. The dataset used for the experiment was split into 70% for the training and 30% for testing. Fig. 4 shows this grouping.

3.2 Evaluation Metrics

In this study, we used the train/test split model evaluation procedure to estimate how well a model will generalize to out-of-sample data and two evaluation metrics to quantify the model's performance. The experiment was performed on a 70% training set to 30% test data split. This was chosen as the train/test split for it is common in the literature. The train/test approach was also considered because of its simplicity, speed, and flexibility.

Two statistical indicators: coefficient of determination ($R^2$) and Root Mean Square Error (RMSE) which are mostly used in the literatures for comparing the performance of different ANN models was used as the evaluation metrics in the present study.

i. The coefficient of determination, or $R^2$, is a measure that provides information about the goodness of fit of a model. It takes values between 0 and 1. The larger this indicator the better the performance of the model. So, the closer the $R^2$ is to unity, the better the estimation ability of the model. This indicator is expressed mathematically as:

$$R^2 = 1 - \frac{\sum(y_a - y_p)^2}{\sum(y_a - y_m)^2}$$  \hspace{1cm} (1)

Where:

- $n$ = number of observations
- $y_a$ = actual value
- $y_p$ = predicted value
- $y_m$ = mean of predicted values

ii. RMSE describes information about the short-term performance by comparing the actual deviation one by one between the estimated and the measured values. It is to some extent the mean absolute deviation error. RMSE is mathematically expressed as:

$$RMSE = \sqrt{\frac{\sum(y_{e,i} - y_{m,i})^2}{n}}$$  \hspace{1cm} (2)

Where:

- $n$ = number of observations
- $y_{e,i}$ = the $i$th estimated value
- $y_{m,i}$ = the $i$th measured value

4 Results and Discussion

In this paper, an ANN was employed to predict daily solar radiation. To demonstrate the merit of the developed ANN model, its prediction performance was evaluated by predicting the daily solar radiation of 30% of the total dataset from 1994-2015. The predicted data using the developed ANN model versus the actual data is as shown in Fig. 5. As can be seen, the ANN is highly non-linear and attempts to follow all the points around the trend.

The performance of the developed ANN model using the evaluation metrics: coefficient of determination ($R^2$) and the root mean square error (RMSE) was then computed. Table 1 summarizes the results obtained using the collected dataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$R^2$</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.78</td>
<td>0.47</td>
</tr>
<tr>
<td>Testing</td>
<td>0.78</td>
<td>0.48</td>
</tr>
</tbody>
</table>

From Table 1, it can be deduced that the model performs above average and hence can be adopted for daily solar radiation prediction.
5 CONCLUSION
A forecasting model based on Artificial Neural Networks was developed to forecast the average daily solar radiation for the North Western region of Nigeria. The developed model is expected to be an assessment tool for predicting and estimating the amount of energy that can be harnessed for solar installation. The performance of the developed ANNs was determined using the coefficient of determination and through an error evaluation metric, the root mean square error. The results obtained showed that the developed forecasting model is reliable and applicable for predicting solar radiation in North Western Nigeria.

REFERENCES